Fundamentals of Genetics

B. D. Singh

FUNDAMENTALS OF GENETICS

B.D. SINGH
Professor
School of Biotechnology
Faculty of Science
Banaras Hindu University
Varanasi

KALYANI PUBLISHERS
LUDHIANA • NEW DELHI • NOIDA (U.P.)
HYDERABAD • CHENNAI • KOLKATA • CUTTACK

FUNDAMENTALS OF GENETICS

B.D. SINGH
Professor
School of Biotechnology
Faculty of Science
Banaras Hindu University
Varanasi

KALYANI PUBLISHERS
LUDHIANA • NEW DELHI • NOIDA (U.P.)
HYDERABAD • CHENNAI • KOLKATA • CUTTACK

KALYANI PUBLISHERS

Head Office
B-I/1292, Rajinder Nagar, **Ludhiana**-141 008

Administration Office
4779, Ansari Road, 23, Daryaganj,
New Delhi-110 002

Works
B-16, Sector-8, **NOIDA** (U.P.)

Branches
1, Mahalakshmi Street, T. Nagar,
Chennai-600 017
110/111, Bharatia Towers, Badambadi,
Cuttack-9 (Orissa)
3-5-1108, Narayanaguda, Hyderabad-500 029
10/2B, Ramanath Mazumdar Street,
Kolkata-700 009

A 0631 04 05 2

© 1990, SINGH, B.D.

First Edition, 1992
Second Edition, 1995
Reprinted, 1997
Third Revised and Enlarged Edition, 2004

Typesetting at TIMES PRINTOGRAPHIC
921, Jatwara Street, Tiraha Behram Khan,
Darya Ganj, New Delhi-110 002

ISBN 81-272-1331-4

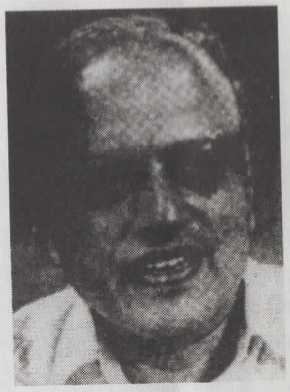
PRINTED IN INDIA
At Plasticote (India), Shahdara Delhi-93
and published by Mrs. Usha Raj Kumar for
Kalyani Publishers, New Delhi - 110 002

The introd
commerci
K

To

Nikhil, Nishith
and
Pushpa

Foreword



Knowledge of Genetics is basic to progress in agriculture, biology and medicine. From 1953 when the double helix model of DNA was discovered, progress in molecular genetics has been phenomenal. This in turn has made sophisticated method of chromosome and genetic engineering available. We are now on the threshold of major advances in our understanding the processes of growth, differentiation and development in living organisms. Unusual opportunities are now open for transferring genes across sexual barriers.

The introduction of Cytoplasmic Inheritance has rendered the commercial exploitation of hybrid vigour possible in many economic plants. Knowledge of population genetics is essential for undertaking improvements of polygenically controlled characters. Prof. B.D. Singh has given a detailed account of different aspects of Genetics and Cytogenetics in this book. We owe him a deep debt of gratitude for the trouble he has taken to compile valuable information and data and present them in a readable manner. The gap in ability to utilise genetic manipulation techniques is the widening between developed and developing countries. We should hence strengthen the teaching of genetics in our Universities. This book is hence a timely contribution.

M.S. SWAMINATHAN
President,
National Academy of Sciences of India

Madras
October 17, 1989

Preface

(3rd Edition)

I am happy to offer the third edition of 'Fundamentals of Genetics'. The science of genetics has seen a drastic change in emphasis: earlier, phenotypic differences were investigated with a view to seek information about genes, while at present DNA sequences are analyzed with the objective of deducing their functional significance. This reversal of approach has been made possible by the powerful tools of recombinant DNA technology, and is expressed in its most imposing magnificence as the human genome project. Any text-book on genetics should accommodate these recent developments, must retain the classical concepts and historical excitements and, even then, keep within the bounds of limited size; this is the kind of challenge that should be avoided, at least, by persons like me who have but limited capabilities.

I have tried to accommodate the recent developments by reducing the coverage of classical areas without, I hope, adversely affecting either the information content or the style of presentation. This made it necessary to deemphasize many treatments and to reduce the level of coverage to suit primarily graduate, *i.e.*, B.Sc. and B.Sc. (Ag.) students, and to provide a solid foundation for post-graduate students. I have also tried to simplify the language as well as the exposition so that students find it easy to learn this fast developing discipline. In order to achieve these objectives, the whole book has been extensively modified; as a matter of fact, only few chapters have remained in their original form.

In this edition, the following seven chapters have been added: (1) Regulation of Cell Cycle (Chapter 6), (2) Gemome Maps (Chapter 26), (3) Recombinant DNA Technology, (Chapter 34), (4) Genetic Transformation of Prokaryotes (Chapter 35), (5) Transgenic Animals (Chapter 36), (6) Transgenic Plants (Chapter 37) and (7) Biosafety (Chapter 38). These chapters attempt to capture some of the recent exciting developments in molecular biology and its practical exploitation for human welfare. In addition, a chronological summary of important developments in genetics is included in the book (Appendices I-IV).

I would like to record my appreciation for the generous support from the teachers and the students of the subject, some of whom were kind enough to point out errors that remained in the earlier editions of the book. I would solicit more such pointers /remarks and suggestions because they provide the stimulus as well as indicate the direction for further improvements.

I would like to express my appreciation for Sri. S.P. Anand, Manager, Kalyani Publishers, New Delhi, and his production staff for their patience and cooperation, and for the excellent execution of this job.

Before I conclude, I would like to record the support and encouragement I have always received from my family, especially, my wife Pushpa and sons Nishith and Nikhil, who have always tried their best to provide me an environment congenial for such work. Their love and thoughtfulness has always been the source of my strength as well as my inspiration for taking up challenging assignments.

Varanasi
April 24, 2003

B.D. SINGH

Preface

(*Ist Edition*)

The science of *Genetics* was born in the spring of 1865 through Gregor Johann Mendel of Brunn, Austria (now Brno, Czechoslovakia) only to be immediately consigned to the deep freeze for the next 35 years; it was rescued and revived in 1900 by three independent workers named Correns, de Vries, and Tschermak. It was to have a troubled childhood, marred by various controversies; a promising adolescence, marked by several exciting discoveries which gave it an identity; and now has an exciting but mindboggling adulthood represented by the promised benefits and the looming dangers of recombinant DNA technology. The degree of dominance this branch of biology has come to exercise on human thoughts and activities is reflected by the two popular terms, 'genetic counselling' and 'biotechnology'. The techniques of recombinant DNA technology have come to be regarded by many as a 'new hope' for industrial and agricultural (and who knows, even social) revolution, but at the same time, they also appear as the greatest threat to our biological world should some crucial regulatory mechanism (either in the recombinant DNA or in the recombinant DNA technologist !) fail to perform its assigned function. Genetics represents one of the greatest achievements of mankind, possibly even greater than those of his space programmes, for it is here he has come close to marshalling the techniques which may well govern the evolutionary destinies of all the biological entities of this world.

It is an enormous task to undertake the writing of a textbook on a subject which begins with the simple concepts of Mendelian laws, and then ramifies into several highly specialized areas, *e.g.*, biochemistry (DNA, gene action, regulation of gene action, etc.), mathematics (quantitative genetics, population genetics), physiology (biochemical genetics), etc. A competent treatment of these diverse areas requires an immense talent and a considerable skill; the talent of comprehending and handling the varied subjects like an expert, and the skill of appearing to have done so even without so doing. In my view, one would require the courage of a daredevil to accept such an ominous challenge. The fact that I accepted the challenge does not necessarily reflect either the courage of my own convictions or a confidence in my own talent and skill; it only shows my gullibility to some well-chosen words of some apparent well-wishers. I was drawn into this quagmire of a project chiefly by two persons. The first of them, Sri Raj Kumar of M/s Kalayani Publishers, set the trap and baited the hook nicely and alluringly. The second person, Sri Ashok Kumar Singh, who I suspect was working for the first one either knowingly or unknowingly, invited me to meet a small group of Research Scholars in IARI. New Delhi; they together succeeded in making me believe that I was well-equipped to achieve

something that, in their view, would be of much help to students. This encouraged me to accept the bait.

What happens to those who rely heavily on other's opinion about themselves would become clear once the reader glances through the pages of this book. But having committed myself to the task, I intend to see it through to the best of my ability. I shall remain indebted to all those who would kindly help me in giving this work an acceptable, preferably, a likeable, shape by making me aware of the areas that need to be reshaped.

I have tried to use a simple language, to build up the various concepts in a rather simple manner and yet to preserve their complex nature, to avoid the unnecessary details for a beginner and yet to keep enough material so as not to make the treatment too preliminary, and so on. It has been my endeavour to present the complex subject as per the needs of the students, and I can only hope that they will find it useful.

Many of my colleagues have contributed in their own way in developing this text; some of these colleagues are, Prof. R.M. Singh, Prof. B. Rai, Dr. A. K. Richharia and Dr. R. Raina. My research scholars, Sri O.P Singh and Sri K. K. Singh, have helped me in one or the other activity related to the project. All my graduate and postgraduate students helped me develop a clearer idea of the various concepts of genetics by posing seemingly innocent but often most devastating questions mostly at the least opportune of the moments. I am grateful to all these and to many others (Sri Y. Singh, Sri A. P. Singh, Sri Ram Sahay and Sri V. N. Shahi) for their help in the preparation of final typescript.

I have been very fortunate in having a wife who understands and helps through the periods of anxiety and preoccupation that go with such projects. My two sons have often, to my immediate irritation, distracted me from what I considered to be my single-minded devotion, only to have me refreshed and reinvigorated so that I could return to the project with a renewed zeal. I feel grateful to my wife, Pushpa, and to my sons, Nishith and Nikhil, for all that they have done for me and for this project.

1990

B.D. SINGH

1

History of Genetics and Cytogenetics

1.1. INTRODUCTION

A keen observer will readily recognise two basic aspects of the biological world. ***First,*** the progeny of an organism develop characteristics similar to that organism. In addition, there are always some similarities between the parents and their progeny. In contrast, the ***second*** aspect of the biological world concerns with the differences among the individuals belonging to a single species. In fact, ordinarily no two human beings are indentical to each other in every respect. *The science of* ***genetics*** *deals with the principles that explain the similarities between parents and their progeny and the differences among individuals of a single species.* In other words, ***genetics is the science of inheritance and variation.***

1.2. DEFINITIONS

Inheritance refers to the transmission of genes governing various characters from parents to their progeny. ***Genes*** are the functional units that govern the development of various characters of an individual. ***Character*** or ***trait*** denotes a morphological, anatomical, physiological, biochemical or behavioural feature of an organism. Some examples of human characters are, eye colour, hair colour, skin colour, presence of hairs on ears, harelip (split upper lip), blood groups etc. The differences among individuals of a single species for a character constitutes the ***variation*** for that character. Since the development of characters is governed by genes, the presence of variation in different characters of an organism indicates the existence variation in the genes governing them. In view of the above, ***genetics may be redefined as the science that deals with the structure, organisation, transmission and function of genes, and the origin of variation in them.***

Cytogenetics *concerns with the study of various aspects of chromosomes and their effects on the development of characters of organisms.* It is universally accepted that *genes are located in chromosomes.* Cytogenetics originated as a result of this realization. It brought together two different branches of biology, namely, cytology and genetics. A study of the structural and functional organisation of different structures within a cell is known as ***cytology***. In cytogenetics the structure and function of chromosomes is studied since they contain genes. *At the gross level, cytogenetics and genetics are distinct from each other. But at the molecular level, the boundary between the two disciplines is not so sharply defined.*

1.3. SPONTANEOUS GENERATION

One of the universally accepted, but almost never stated, facts of genetics and cytogenetics is as follows: *an organism originates from a preexisting organism of similar kind.* Thus an organism does not arise or originate directly from nonliving

matter. But till the middle of the nineteenth century, many biologists believed that some primitive organisms originated from nonliving matter, *e.g.*, decaying organic matter. Such a *de novo* (afresh) origin of living organisms from nonliving materials is known as ***spontaneous generation.*** Aristotle (384-322 B.C.), the greatest biologist of early days, believed that not only plants but animals like fleas, mosquitoes and snails originated from decaying organic matter. This belief was held by practically all the biologists until the middle of the eighteenth century. As a result, when Redi (1621-1697) and Spallanzani (1729-1799) presented experimental evidence against spontaneous generation, every conceivable argument against these experiments was put forth.

Redi placed pieces of meat in clean vessels and covered these vessels with clean fine cloth. This prevented adult flies from reaching (and laying eggs on) the meat. As a result, larvae of fly did not originate from the rotting meat, as they did when the meat was left open. This clearly demonstrated that sitting of adult flies on meat pieces was essential for the origin of fly larvae from rotting meat.

Spallanzani presented evidence against the spontaneous origin of micro-organisms. He filled liquid organic matter like bean soup in flasks, boiled the organic matter in the flasks for a sufficiently long time and sealed the flasks by melting their neck. Examination of the contents of these flasks even after several days revealed that there were no microorganisms in the organic matter. This clearly showed that microbes did not arise spontaneously. But many biologists argued that the flasks contained boiled or 'tainted' air, which had lost its 'vital force'. it was further suggested that if the flasks contained fresh air there would have been spontaneous generation of microbes.

The controversy between the supporters and the opponents of spontaneous generation was finally resolved by Lois Pasteur (1822-1895) and Tyndall (1820-1893). Working independently, they provided incontrovertible evidence against spontaneous generation. These scientists boiled organic matter in narrow-mouth flasks for sufficiently long time, and then tightly plugged the mouth of these flasks with clean cotton. Cotton plugs permitted the entry of fresh air into the flasks but kept the microbes out of them. The organic matter in these flasks did not decay, and no microbes could be recovered from them even after several days. Thus *Pasteur and Tyndall conclusively proved that microbes do not originate spontaneously from organic matter.*

1.4. REPRODUCTION IN PLANTS AND ANIMALS

As early as 700 B.C., Babylonians and Assyrians were aware of the fact that female palm trees produce fruits only when their flowers were dusted with pollen. But a systematic effort to understand the significance of pollination in plants began much later.

Eggs of birds like poultry and their role in reproduction were long known. But the eggs of other animals like mammals were discovered much later after the light miscrope was developed. In 1673, de Graaf proved that the tiny vesicles on the surface of mammalian ovaries were eggs. He also *noted that progeny had the characteristics of both their parents. On this basis, he suggested that both the parents contribute to heredity.* About the same time, Malpighi in Italy and Swammerdam in Holland identified the eggs of insects. In 1674, Leeuwenhoek and his student Hamm studied the sperms of many animals. in 1680, Leeuwenhoek noted

that in sexual reproduction sperms are associated with eggs, but he did not speculate on whether they fused or not.

The roles of egg and sperm in sexual reproduction were set in the right perspective by a series of important discoveries. In 1785, Spallanzani produced puppies through artificial insemination in dogs. By filtering the seminal fluids of frogs and toads, he also proved that the presence of sperm was essential for the development of progeny. *A conclusive proof for the fusion of sperm and egg was presented by O. Hertwig in 1875 in sea urchin; he observed the fusion of sperm and egg nuclei.*

Grew described stamens in flowers and suggested that they were the male reproductive parts. *In 1694, Camerarious showed that maize plants did not set seed when pollen was not applied to their pistils.* But most eighteenth century biologists found it hard to believe that plants reproduced sexually in the same manner as animals. *The first artificial hybrids in plants were produced by Thomas Fairchild in 1717 when he crossed sweet William with carnation.* Subsequently, many scientists, *e.g.*, Linnaeus (1707-1778), Koelreuter (1733-1806), Gaertner (1772-1850) etc., produced artificial hybrids in plants. They noted equal contributions by both the pollen (male) as well as the seed (female) parents to the characters of progeny. In 1823, Amici described the entry of pollen tubes into the ovary. *Strausberger described fertilization in angioperms and noted the fusion of sperm and egg nuclei in 1884.*

1.5. PREFORMATION

It was long known that reproduction in animals involved the union (copulation) of males with females. However, the contribution of male and female parents to the characters of progeny, *i.e.*, heredity, was a matter of great debate. Aristotle believed that the progeny derived their form from semen, while the substance (nutrition) was provided by the females. This concept was further developed by the scientists, called ***animalculists,*** of the seventeenth and eighteenth centuries. On the other hand, Harvey (1578-1657) speculated that the progeny developed from eggs and that semen only produced a stimulatory effect (called ***aura seminalis)***. This suggestion was further elaborated by the scientists called ***ovists.***

After the discovery of eggs and sperms, many biologists began a detailed study of these gametes. In 1694, someone examining human sperms thought that he had seen a miniature human figure, called ***'homunculus',*** inside the sperm. This report was enthusiastically supported by Swammerdam and Bonnet (1720-1793). Swammerdam calimed that homunculi were present in sperms and that only the sperms contributed to heredity. A group of biologists (*animalsculists*) supported this view. Bonnet believed that homunculi were located in eggs and that only the eggs contributed to heredity; this idea was supported and elaborated by another group of biologists (*ovists*). According to these scientists, a miniature, but complete in all respects, human being was already present in eggs/sperms, *i.e.*, new human beings were already present in the gametes. Thus the development of progeny involved only the growth of this preformed human being. This hypothetical phenomenon is termed as ***preformation*** and those who held this belief are called ***preformationists.*** Even Leeuwenhoek was a preformationist (animalculist). However, many biologists working with improved microscopes failed to see homunculi either in eggs or sperms. As a result, the idea of preformation was replaced by a more acceptable view of epigenesis.

1.6. EPIGENESIS

The greatest setback to the idea of preformation was provided by studies on embryo development. These studies revealed that the different organs and tissues of adult plants and animals developed from the uniform embryonic tissues. Clearly, the organs and tissues of an organism did not arise from a mere growth of those of the miniature homunculi present in eggs/sperms. These organs and tissues arose through a process of differentiation from the homogeneous embryonic tissues. In the early stages of embryonic development, however, embryonic tissues are composed of similar cells; these cells progressively differentiate into the adult tissues and organs. This concept is known as ***epigenesis***; it was originally proposed by Wolff (1738-1794). However, Wolff believed that adult tissues originated completely *de novo*. But his famous successor, von Baer (1792-1876), proposed that *adult tissues did not originate* ***de novo****; they developed through a sequential modification (differentiation) of the embryonic tissues themselves. This concept is the universally accepted view of origin of the adult organs and tissues from zygotes.*

HOMUNCULUS (PRESENT WITHIN OVA/SPERMS)

GROWTH/EXPANSION OF THE PREEXISTING HOMUNCULUS → EMBRYO/BABY

PREFORMATION

OVUM + SPERM → ZYGOTE → CELL DIVISION → MASS OF IDENTICAL CELLS → DIFFERENTIATION → TISSUES/ORGANS OF THE EMBRYO

EPIGENESIS

Fig. 1.1. A comparison of the various events in the development of human embryo according to ***preformation*** and ***epigenesis***. Note that according to preformation, a fully differentiated human being is already present within the gametes. On the other hand, the embryo differentiates from a mass of identical cells according to the theory of epigenesis.

1.7. INHERITANCE OF ACQUIRED CHARACTERS

In an attempt to explain the origin of differences in the characteristics of different species, that is, variation, *Lamarck (1744-1829) proposed that characters acquired by individuals of one generation are transmitted to the next generation.* The idea of the ***inheritance of acquired characters,*** popularly known as ***lamarckism***, was originally suggested by Hippocrates about 400 B.C. An ***'acquired character'*** refers to the intensity and form of a character of an individual that develops primarily due to the effects of environmental factors. Thus according to the idea of inheritance of acquired characters, if a person develops strong muscles by exercise, his children will inherit this characteristic. Conversely, if a person became weak due to disease or some other reason, his children would also be weak. It was an attractive hypothesis to explain the development of distinctive characteristics of various species during evolution. This hypothesis had several notable supporters including Charles Darwin (1809-1882). More recently, Lysenko and his supporters in Russia attempted to collect evidence in favour of this hypothesis.

Evolution

- Charles Darwin (1859) published 'Origin of Species'.
- Darwin proposed that evolution occurs due to 'natural selection' resulting in the' 'survival of the fittest' (called *Darwinism*).
- Modern view: natural selection has a quantitative effect on survival; it rarely eliminates a genotype.

1.8. ORGANIC EVOLUTION

Charles Darwin published his classic *Origin of Species* in 1859 based on the material collected by him from Malay Archipelago and the other places he visited on board Beagle. Darwin proposed the ***theory of evolution*** according to which two or more different species evolved or differentiated from a single parent species primarily due to the effects of natural selection on the variation present in the species. According to Darwin, natural selection allows only the fittest individuals of a species to survive; this is known as ***survival of the fittest*** *due to natural selection.* Today, *the concept of evolution through natural* selection is universally accepted. *But it is also recognised that the effect of natural* selection is quantitative and *gradual rather than qualitative and abrupt. As a result, both the fittest and the less fit types survive, and the less fit types are not eliminated. However, the fittest types leave behind a greater number of progeny than the less fit types; this difference is the effect of* ***natural selection.***

Darwin accepted the hypothesis of inheritance of acquired characters as one of the reasons for origin of variation in characters of a species. At the same time, Darwin attempted to provide a physical basis to this hypothesis. He proposed an elaborate scheme, called ***pangeneis***, which runs as follows. Each organ of an individual produces very small, almost invisible identical copy of itself called ***gemmule*** or ***pangene.*** Gemmules from all the organs of an individual are transported into the gonads. Here, the gemmules are distributed into the gametes so that each gamete receives a gemmule for each organ of the individual. When male and female gametes unite during fertilization, the zygote receives gemmules from both male and female parents. During development, gemmules from the two parents migrate into the respective organs and determine the development of these organs. If an organ of an individual is modified in some way, the gemmule for that organ will also be modified accordingly. This modified gemmule will be transmitted through the gametes to its progeny and will produce a similar modification in the corresponding organ of the progeny.

The idea of ***pangenesis*** *is of only historical interest without any scientific validity. It was negated, soon after its proposal, by Weismann* (1834-1914) *through a series of brilliant experiments on rats.*

1.9. THE GERMPLASM THEORY

In order to test the validity of the hypothesis of inheritance of acquired characters (due to Lamarck) and of pangenesis (due to Drawin), Weismann cut off the tails of a group of experimental mice for 22 successive generations. According to the hypotheses of inheritance of acquired characters and pangenesis, the tailless condition of mice so produced should be transmitted to their progeny. But *Weismann did not obtain a single tailless progeny even after consistent removal of tails for* 22 *generations. This conclusively disproved the ideas of inheritance of acquired characters and pangenesis.*

Germplasm theory (Weismann 1834 - 1914)

- Body of an individual consists of :
  - (*i*) *Somatoplasm* (not involved directly in sexual reproduction) and
  - (*ii*) *Germplasm* (produces gemetes).
- *Changes only in grmplasm inherited.*
- Validity proved by Castle and Phillips (1909) by ovary transplant in guinea pigs.
- Replaced by the concepts of genotype and phenotype due to Johannsen (1903, 1909).

Weismann proposed the ***germplasm theory*** to explain these and other findings. According to this theory, *the body of an individual is composed of two distinct types of tissues*: (1) *somatoplasm and* (2) *germplasm.* ***Somatoplasm*** consists of all the body tissues, which do not contribute to sexual reproduction. Therefore, *changes occurring in somatoplasm cannot be transmitted to the next generation.* The ***germplasm,*** on the other hand, produces gametes that are the basis of sexual reproduction. *The changes taking place in the germplasm only are, therefore, transmitted to the next generation.*

In an animal, gonads represent germplasm, while the remaining body represents somatoplasm. Germplasm and somatoplasm do not show any fundamental differences, except for their function, *i.e.*, production of gametes. In addition, microbes and other single-celled organisms do not have distinct somatoplasms and germplasms.

The fact that characters of somatoplasm are not inherited was conclusively demonstrated by Castle and Phillips in 1909. In guinea pigs, albino coat colour is produced by a single recessive gene; its dominant allele produces black fur colour. Castle and Phillips removed the ovaries of an albino guinea pig and transplanted into this animal an ovary from a black guinea pig. This female was mated with an albino male. *If pangenesis/inheritance of acquired characters were valid, all the progeny from this mating must be albino. On the contrary, all the progeny from this mating had black fur. This can be explained only if it is assumed that the somatoplasm or the body tissues of albino female had no effect on the function of the gonads transplanted into it from the black female. These findings, together with those of Weismann, conclusively demonstrate the separateness of germplasm and somatoplasm and disprove the hypotheses of pangenesis and inhertance of acquired characters.*

1.10. HYBRIDIZATION IN PLANTS

Thomas Fairchild produced the first artificial hybrid in plants when he crossed caranation and sweet William in 1717; this hybrid was known as *Fairchild' mule.* Subsequently, a number of biologists, *e.g.*, Joseph Koelreuter, John Knight, John Goss, Sargret, Gaertner, Darwin, Naudin, Linnaeus etc., worked on plant hybridization. *Koelreuter* made extensive crosses in *Nicotiana* between 1760 and 1766. He *noted the uniformity and heterosis in* F_1 *and the appearance of a larger variation in* F_2. *He also concluded that both male and female parents make equal contributions to the characteristics of progeny.*

John Knight (1759-1835) developed commercial varieties of apple, pear, apricot, grapes etc. through hybridization. The results obtained by Gaertner (1772-1850), Naudin (1815-1909), Darwin and others agreed with those of Koelreuter. They also presented evidence for dominance, although the phenomenon was named by Mendel sometime later. *The observations of these scientists clearly showed that some characters of the* F_1 *progeny resembled those of one parent, some others were similar to those of the other parent, while some others were intermediate between those of the two parents. The observations and experiments of these workers provided the basis for the historical experiments in plant hybridization by Mendel.*

1.11. LAWS OF MENDEL

It was a widely held belief among the biologists during and before the time of Mendel that the genetic material, *i.e.*, the medium for inheritance, was a liquid, possibly blood. Consequently, the genetic material from the two parents was thought to blend in the progeny so that the characters of progeny would be intermediate between those of the two parents. This idea of ***blending inheritance*** was adequate to explain the intermediate position of the F_1 between the two parents. But when the hybrid resembled one or the other parent, it was explained on the basis of unequal contribution by the two parents. Sometimes, such irrelevant factors as the direction of wind at the time of fertilization were also invoked to explain the similarity of a hybrid with one of its parents for a character.

Mendel's Laws

- Mendel (1866) showed that genes :
  - (*i*) are particulate (*disproves the idea of blending inheritance*).
  - (*ii*) segregated during gamete formation (*law of segregation*), and
  - (*iii*) two different genes show independent segregation (*law of independent assortment*).
- *Science of* ***Genetics*** *is borne.*

However, *blending inheritance could not explain the appearance of a much larger variation in* F_2 *as compared to that in* F_1. According to blending inheritance, all the gametes produced by F_1 should be identical for the character in question. Therefore, F_2 progeny should be uniformly intermediate like F_1. This can be explained more clearly with the help of an analogy. Let us suppose that there are two beakers; one beaker has 100 ml of a sugar solution, while the other has 100 ml water. When both the sugar solution and the water are poured into a third beaker

they mix together; the sweetness of this solution will be less than that of the original solution but greater than that of water. If the new solution is distributed into a number of beakers, they will all contain solution of the same sweetness, and we cannot expect the sweetness of solutions in these beakers to show a detectable variation. Similarly, F_2 progeny should be as uniform as F_1 if the idea of blending inheritance were valid.

Mendel conducted a series of experiments on hybridization in peas; the findings from these experiments were published in 1866. From these studies, ***Mendel concluded that the unit of inheritance, which he referred to as* 'factor' *and we now call* 'gene', *is particulate and not liquid. The two alleles of a gene do not affect each other in any way,*** what to say of blending with each other. ***Mendel proposed two universally accepted laws of inheritance, the law of segregation and the law of independent assortment, and explained the dominance of characters. The two laws of Mendel, especially the law of segregation, are the basis of the science of genetics, and Mendel is aptly known as the* 'father of genetics'.**

But the contemporaries of Mendel neither understood nor appreciated the significance of his findings. *His paper lay forgotten in the stacks of many libraries till* 1900 *when it was rediscovered independently by Correns, de Vries and Tschermack.* These scientists themselves had obtained results similar to those of Mendel. The rediscovery of Mendel's paper initiated a spurt of research activity in the field of genetics. This has resulted in the phenomenal development of this branch of knowledge in the brief span of 102 years.

1.12. DISCOVERY OF CELL

Janssen and Janssen were the first to construct a properly working compound microscope in 1590. Leeuwehoek made notable improvements in the lenses of compound microscope. In 1665, Robert Hooke cut a thin section of a cork and observed it under a microscope. He found that the cork was made up of a large number of empty chambers similar to those of a beehive. Hooke referred to these chambers as ***cells.*** In fact, the cork cells were dead; living cells are never empty but they are filled with cytoplasm.

1.13. THE CELL THEORY

In 1824, Dutrochet concluded that all biological tissues are composed of cells and that these cells are attached with each other due to some adhesive compounds. A few years later, *in* 1838, *Schlieden published his studies on the structure of plant tissues and concluded that all plant tissues were made up of cells. In the next year, in* 1839, *Schwan concluded the same thing for animal tissues on the basis of his studies.* Thus by 1839, conclusive evidence was presented to prove that cell is the structural unit in both plants and animals. It was further postulated that cell is also the functional unit of living organisms.

These conclusions were organised in the form of ***cell theory.*** *According to this theory, cell is the structural as well as the functional unit of all living organisms.* This theory is credited jointly to Schlieden and Schwan and is universally accepted. But one notable exception to cell theory are viruses, which do not possess a typical cellular organisation. In fact, viruses contain only nucleic acids and proteins, and many biologists treat them on the borderline of living beings.

The living organisms are either unicellular, *e.g.*, bacteria, protozoa, some algae, yeast etc., or multicellular, *e.g.*, most fungi and all other plants, and a vast majority of animals. In single-celled organisms, the cells are self-sufficient units of both structure and function. But in multicellular organisms, the cells, as a rule, differentiate to various degrees to form different tissues and organs. Differentiation leads to changes in both the structures and functions of cells as they become increasingly specialized. Therefore, many cells lose the typical structure for a cell. Some cells even lose their nucleus, *e.g.*, red blood cells, companion cells of phloem, or even their entire cytoplasm, *e.g.*, tracheary elements of xylem etc. In all multicellular organisms, every cell changes in structure in a very regular and specific manner so that usually normal organisms develop. Therefore, these cells still function as structural units in the organisation of the organism.

Similarly, each cell in the body of a normal organism differentiates to perform a specific function or group of functions. This differentiation is again highly regulated along specific patterns and is extremely regular. The normal functions of an organism arise directly from functions of the divergent cells if its body, which have differentiated in highly specific and regulated patterns. These cells, therefore, do work as functional units, although each cell type performs only specific function(s).

Thus we can see that the concept of cell theory is applicable to all the living organisms, whether unicellular or multicellular, except viruses. In 1846, Nageli provided the logical extention of cell theory: he postulated that all cells originate from preexisting cells. Virchow, in 1853, presented an elaborate version of this hypothesis along with considerable evidence in its support; it is referred to as ***cell lineage theory*** and is credited to Virchow.

1.14. NUCLEUS AND CHROMOSOME

Leeuwenhoek was the first to observe the nucleus of red blood corpuscles in 1674. But he did not recognise it as a separate organelle and could not appreciate its significance. *Nucleus was discovered and named by Brown* in the cells of *Tradescantia* in 1838. In 1871, Meischer developed the technique for isolation of nuclei from tissues; he also extracted nucleoproteins and nuclein from nuclei. Later, in 1889, Altman renamed the nuclein as nucleic acid, which is the universally accepted term today. *Flemming proposed the term mitosis in 1882; he also described longitudinal division of chromosomes during cell division.* Flemming further suggested a relationship between chromatin and nucleic acid. In 1888, Strausberger demonstrated that during the formation of male and female gametes the chromosome number is reduced to half. *This type of cell division was referred to as meiosis by Farmer and Moore in* 1905. *The term chromosome was first used by Waldeyer in 1888. In 1902, McClung described the X chromosome of Hemiptera.*

1.15. CYTOGENETICS

In 1903, Walter Sutton explained the significance of reduction division, and proposed the *chromosome theory of heredity.* Sutton analysed the experimental findings in the fields of cytology and genetics and concluded that both genes and chromosomes exhibit identical behaviour during cell division, fertilization, development and sexual reproduction. *He proposed that genes were located in chromosomes and showed that this adequately explained the laws of segregation and independent assortment of Mendel.* In the same year, Boveri concluded that

chromosomes played an important role in development and that genes were present in chromosomes. *Therefore, Sutton and Boveri are jointly credited with the* ***chromosome theory of inheritance.*** This marked the birth of cytogenetics.

1.16. CLASSICAL GENETICS

The terms ***gene, genotype*** and ***phenotype*** were used by Johannsen in 1903. Two year later, in 1905, *Bateson introduced the term* ***genetics***. In the same year, Bateson observed linkage between two traits in pea, but he could not provide the correct explanation for the phenomenon. In 1905, Bateson and Punnet published an account of a modification of the dihybrid F_2 ratio due to gene interaction. In 1903, Johannsen published his findings on selection for seed size in the Princess variety of rajma (*Phaseolus vulgaris*). Johannsen proposed the ***pureline theory,*** which takes into account variations arising due to genotype and environment. A ***pureline*** is the progeny of a single homozygous self-pollinated plant. He also gave the concepts of ***genotype*** (the genetic makeup of an individual) and ***phenotype*** (the external appearance of an individual). *The concepts of genotype and phenotype have replaced those of germplasm and somatoplasm proposed by Weismann as they are more precise and informative than the latter.*

The idea of ***multiple factors*** was advanced by Yule in 1906 in order to explain the inheritance of quantitative or metric characters. *According to this hypothesis, several genes govern the development of a single quantitative character. Each of these genes has a small effect, and effects of these genes are additive in nature. The existence of multiple factors was demonstrated by Nilsson-Ehle in 1909* through his studies on the inheritance of seed colour in wheats and oats. The ***multiple factor hypothesis*** has provided the genetic explanation for inheritance of quantitative characters. This hypothesis has replaced the earlier, seemingly attractive but scientifically impossible, hypothesis of blending inheritance.

In 1910, Morgan published a short paper containing his findings on the inheritance of *white* eye (*w*) gene of *Drosophila melanogaster.* He explained the unusual inheritance of this character by assuming that the *white* eye (*w*) gene was located in the *X* chromosome of *Drosophila. This was the first conclusive evidence in favour of the chromosome theory of heredity and the* w *gene of* Drosophila *became the first gene to be located onto a specific chromosome. This study demonstrated* ***sex linkage*** *and indicated the existence of linkage between genes,* which became clearly evident from the subsequent studies by Morgan and his group. Morgan and coworkers developed the technique of chromosome mapping through linkage studies, and prepared the first ever linkage map of *Drosophila. In 1934, Morgan was awarded the Nobel Prize for these studies.*

Muller demonstrated the mutagenic effects of X-rays in Drosophila *in* 1927. This provided a tool for inducing changes in genes and raised the possibility of desired genetic manipulations in organisms. Sometimes later in the same year, Stadler reported the induction of mutations in barley and maize by X- and gamma-rays; this confirmed the findings of Muller. *Muller was awarded the Nobel Prize in 1946 for this work.*

1.17. BIOCHEMICAL GENETICS

A British physician named Garrod published a book, '*Inborn Errors in Metabolism*', *in 1909* in which he described some hereditary diseases in man. *He concluded that*

persons suffering from a hereditary disease lacked a specific enzyme resulting in abnormal metabolism in them. Conclusive evidnce supporting this hypothesis was presented by Beadle and Tatum in 1941 from their studies on ***biochemical mutants*** (mutants unable to synthesise, and hence requiring a specific biochemical for normal growth) of the common bread mold *Neurospora crassa. They proposed that each gene produces a specific enzyme, which catalyses a specific biochemical reaction; this is commonly known as* ***one-gene-one-enzyme hypothesis.*** *This study founded a new branch of genetics,* ***biochemical genetics.*** It also provided an indication of the manner in which a gene governs the development of a character. *Beadle and Tatum (along with Lederberg for his work on bacterial conjugation with Tatum) were awarded the Nobel Prize in* 1958.

1.18. MOLECULAR GENETICS

In 1944, Avery, MacLeod and McCarty demonstrated that DNA (deoxyribonucleic acid) was the genetic material. *The* ***double-helix model of DNA*** *structure was proposed by Watson and Crick in* 1953; *they, along with Wilkins, were awarded the Nobel Prize in 1962 for this work.* The double helix model of DNA is universally accepted and the experimental data accumulated so far agree with this model. These discoveries laid the foundation of a new branch of genetics called ***molecular genetics***.

In 1950, McClintock proposed an unusual genetic element (gene) called dissociator (Ds) *in maize. This gene has the ability into change its position within a chromosome and also to move from one chromosome to another.* This is contrary to the general concept of a gene as each gene is believed and experimentally *shown to occupy a specific and fixed position within a chromosome*: a gene does not change its location unless there is a corresponding change in the chromosome structure. For this reason, *the term locus is used as a synonym for gene;* ***locus*** means a fixed position. It is now known that the vast majority of genes occupy fixed positions in the respective chromosomes. *But some genes are capable of changing their position; such genes are known as* ***transposon, transposable elements*** *or, more popularly,* ***jumping genes.*** McClinktock had proposed the concept of jumping genes much ahead of her time. The significance of this concept was appreciated only in the late seventies, and *she was awarded the Nobel Prize in* 1983.

In 1961, Nirenberg and coworkers began to decipher the genetic code through a series of brilliant experiments. ***Genetics code*** is often referred to as ***triplet code*** since the code for one amino acid (***codon***) has three nucleotides. *They were the first to demonstrate that three uracils (UUU) in messenger RNA (mRNA) code for a molecule of phenyl alanine.* Subsequently, Nirenberg and coworkers determined the codes for other amino acids; today, the meanings of all the possible 64 codons are known. *Nirenberg was awarded the Nobel Prize in* 1968 *for this work.*

In the year 1961, two French scientists, Jacob and Monod, proposed the ***operon concept*** of gene regulation. *According to this concept, the activity of a group of structural genes (genes producing polypeptides/enzymes) is controlled by three genes known as* ***regulator, operator*** *and* ***promoter.*** *The operon concept is universally accepted, and Jacob and Monod were awarded the Nobel Prize in* 1965.

Benzer published the results from his studies on the fine structure of r^{II} locus of T_4 bacteriophage about the same time. He proposed three different units in genetic organisation: (1) ***cistron,*** the unit of gene function or production of polypeptide; (2) ***muton,*** the unit of mutation or the smallest part of a gene in which mutation can

occur; and (3) ***recon***, the unit of recombination or the smallest part of gene within which recombination *can not* take place. These terms are seldom used, and *gene is now used in the same sense as cistron.*

Ochoa achieved the *in vitro* synthesis of RNA (ribonucleic acid) in 1956. In the same year, Kornberg developed the technique for *in vitro* synthesis of DNA. *Ochoa and Kornberg were jointly awarded the Nobel Prize in 1959 for these achievements. Dr. Hargovind Khorana, a scientist of Indian origin, was awarded the Nobel Prize in 1968 for his work on genetic code and* in vitro *synthesis of DNA.* In 1970, Khorana and associates synthesized a known gene coding for alanyl-transfer RNA *in vitro;* subsequently they demonstrated that this *in vitro*-synthesized gene functioned normally in transcription, *i.e.*, functioned in the synthesis of RNA as a template or mould.

1.19. RECOMBINANT DNA TECHNOLOGY

A brief survey of the developments in the field of genetics between 1900 and 1984, shows that this branch of biology has come a long way from the point of its initiation. It has almost provided a complete understanding of the issues it set out to investigate, *i.e.*, the phenomena of heredity and variation. Recently, techniques for the production of recombinant DNA have been developed.

Recombinant DNA is produced by joining together two or more DNA segments usually originating from different organisms. This is achieved by using specific restriction enzymes for cutting the DNA molecules into suitable fragments and then joining these fragments by using a DNA ligase. Restriction enzymes are a kind of endonuclease. Endonuelease was first isolated in 1970 for which Smith, Nathans and Arber were awarded the Nobel Prize in 1978. The DNA molecule to be cloned is inserted into a DNA molecule called ***vector***, which has the ability to propagate itself if placed into a suitable organism (***host***). Ordinarily, a vector molecule containing a DNA fragment that is to be cloned, is known as ***recombinant DNA***. The technology of production of recombinant DNA molecules is called ***recombinant DNA technology***, while multiplication of the recombinant DNA molecules in a suitable host is known as ***gene*** or ***DNA cloning*** or ***genetic engineering***. *But often recombinant DNA technology and gene cloning are used as synonyms.* The first successful genetic engineering experiments in bacteria were performed in 1973.

Recombinant DNA technology or gene cloning may have one of the following three objectives: (1) to obtain a large number of copies of a DNA sequence/segment, (2) to recover large quantities of the protein encoded by the cloned DNA sequence, and (3) to integrate the DNA sequence into the genome of the target organism (this produces ***transgenic organisms***).

1.20. FORWARD AND REVERSE GENETICS

The science of genetics began with the objective of understanding the principles governing and the mechanisms of heredity and variation. *Its starting point was phenotype*: differences in phenotypes were subjected to increasingly sophisticated and penetrating analyses, and the various kinds of spontaneous and induced mutants provided the most vital materials for these studies. These quests culminated in the development of the various concepts of genetics, including those of cistron, identification of the genetic material, understanding of the genetic language itself and even isolation of several genes. These studies and approaches constitute ***forward***

genetics as one begins with the phenotype of an organism and proceeds sequentially to identify the gene or DNA sequence responsible for this phenotype (Fig. 1.2).

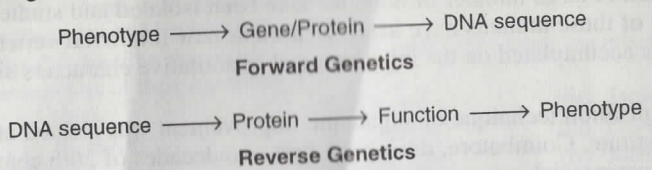


Fig. 1.2. Schematic representation of forward and reverse genetics.

But recombinant DNA technology has made it possible to isolate any DNA segment from the genome of an organism and to clone and express it in a suitable host organism, *e.g.*, bacteria, yeast, *Xenopus* occyte or even a plant/animal. From such a study, one can determine the function performed by a DNA sequence, and then assess the most likely role of this function in the development/organization/function of the organism. This approach of genetic analysis is most aptly described as ***reverse genetics*** since here we begin with a DNA sequence of unknown function, and then proceed to determine its role in the development of phenotype (Fig. 1.2).

Thus in *forward genetics* investigation proceeds from phenotype to DNA sequence, while in *reverse genetics* it progresseses from DNA sequence to phenotype.

Forward and Reverse Genetics

- Study begins with phenotypic variation. progresses to identification and even isolation of gene (***forward genetics***).
- A DNA segment is first isolated, its sequence/function is then determined and its role in phenotype development surmised (***reverse genetics***).

1.21. EPILOGUE

With the help of recombinant DNA technology man is becoming capable of effecting desired modifications in genes and of synthesising any desired gene. These techniques may provide man the power to accelerate the rate and even change the direction of organic evolution. This is indeed the greatest development in biology, which may give man an unlimited power to control his own destiny. However, there is a disturbing feature of this great development. Recombinant DNA technology holds a great promise for human welfare, but it can also be turned against humanity by employing it to develop highly destructive antibiotic-resistant pathogens to be used in a biological warfare. It will be a sad day for humanity if this powerful tool was put to such a destructive use. Let us hope that man will rise to the challenge of the situation and prove himself to be indeed the most intelligent of all the creatures on earth. We pray that he may not use his knowledge of the secrets of life to annihilate life itself.

1.22. GENETICS IN INDIA

Genetic research in India began in 1906 when Imperial Agricultural Research Institute began to function at Pusa, Bihar. Subsequently, other research institutes were established. But for a long time, research at these centres was confined to the

development of new improved varieties of different crops and genetic analyses of various characters of crop plants. During 1930s, work was initiated on mutagenesis of crop plants. A large number of mutants have been isolated and studied in several crops: some of these mutants have been released as new improved varieties. A large literature has accumulated on the inheritance of quantitative characters since the late 1950s.

The noblisation technique of sugarcane improvement was refined at Sugarcane Breeding Institute, Coimbatore, during the first two decades of 20th century. In this technique, commercial sugarcane varieties are developed by crossing noble canes (*Saccharum officinarum*) with *Saccharum* spp., *e.g., S. barberi, S. spontaneum. S. robustum* etc. Today, commercial sugarcane varieties all over the world are developed through the noblisation technique; such varieties are known as ***noblised canes.*** In 1916, C.A. Barber produced an intergeneric hybrid between *S. officinarum* × *Narenga porphyrocoma.* Subsequently, during 1930s, a hybrid between *Saccharum* and *Zea mays* (maize) was produced. Investigations during 1950s and 1960s revealed that in the interspecific crosses among *Saccharum* spp., *e.g., S. officinarum* × *S. spontaneum, S. officinatum* × *S. barberi* etc., often unreduced (2*n*) gametes of *S. officinarum* function and give rise to interspecific hybrids. The same situation was detected in the intergeneric crosses involving *S. officinarum.* (Conventionally, the female parent of a cross is written first on the left side of the symbol '×', while the male parent is written on the right side of '×'.)

Cytogenetic studies on plants and animals began in early 1930s. Extensive studies on chromosome number and morphology have been carried out in all the groups of plants (bryophyta, pteridophyta and flowering plants) and animals. These studies have been quite helpful in cytotaxonomic (taxonomic classification with the aid of cytogenetic data) and evolutionary studies. So far, more than 25% of the Indian flora is estimated to have been cytogenetically analysed. During early sixties, accessory or B chromosomes were demonstrated in woody trees, while thus far they were regarded to be confined to herbaceous plant species only.

Genetic studies on lower organisms were quite late to start during 1950s. Several workers have concentrated on the isolation and characterization of morphological as well as biochemical mutants of these organisms. Using such mutants, gene transfer was demonstrated to occur in some blue-green algae by some Indian scientists during 1960s. These findings have since been confirmed by other workers as well.

Investigation on molecular and developmental genetics were initiated in some laboratories during late 1970s and early 1980s. Advanced and sensitive experimental techniques are being used to collect information on the molecular organisation of eukaryotic chromosomes and other molecular aspects of cytogenetics. Some laboratories have also initiated work on recombinant DNA in microorganisms, animals and plants. The Government of India are promoting an ambitious research and training programme in biotechnology, some aspects of which utilize advanced molecular genetic techniques. Genetically engineered lines of some crops have been produced and are being evaluated before release for commercial cultivation.

QUESTIONS

1. Define genetics and cytogenetics. Briefly describe the research findings related to genetics before Mendel.

2. Briefly describe the research work on cytology and genetics that led to the origin of cytogenetics.
3. Give a brief description of the important discoveries in genetics after the rediscovery of Mendel's paper.
4. Briefly describe the discoveries made by the following scientists with a critical assessment of their significance:

(*i*) Redi	(*ii*) Pasteur	(*iii*) Swammerdam
(*iv*) W. Wolff	(*v*) Lamarck	(*vi*) Darwin
(*vii*) Weismann	(*viii*) Meischer	(*ix*) Mendel
(*x*) Koelreuter	(*xi*) Corens, de Vries and Tschermak	(*xii*) Schleiden and Schwan
(*xiii*) Brown	(*xv*) Morgan	(*xvi*) Watson and Crick
(*xiv*) Flemming	(*xviii*) Hertwig and Strausberger.	
(*xvii*) Sutton and Bovery		

5. Give a list of the scientists awarded, Nobel Prize for their work in genetics along with the year of award. Briefly describe the research work for which they were awarded the prize.

2

Cell Ultrastructure

2.1. INTRODUCTION

Cell is the structural and functional unit of all living organisms, except viruses. Cells show a large variation, but all the cells belonging to a single tissue are generally similar in size and shape. Cells may be spherical, cylindrical, rod-shaped, hexagonal cylindrical or of irregular shape. Red blood corpuscles (RBCs) have a depressed central area and resemble a 'balushahi' in shape. The size of cells ranges from 0.2 to 5.0 μ (μm, microns or micrometers) in bacteria to up to 75 mm in the case of ostrich eggs (without shell). But most of the cells are cylinders of 15-30 μ in diameter. The number of cells in an individual may range from one (bacteria, protozoa etc.) through 6×10^{12} (human beings) to even a larger number in animals like whales and elephants. The following structures are visible in a cell with the aid of a light microscope: cell wall, cytoplasm, nucleus, chloroplast and mitochondria. However, studies with electron microscope have revealed many more structures within an eukaryotic cell. The various structures seen in plant cells are, (1) cell wall, (2) plasma lemma, (3) endoplasmic reticulum (ER), (4) ribosomes, (5) Golgi bodies, (6) lysosomes, (7) spherosomes, (8) chloroplasts, (9) mitochondria, etc. and (10) nucleus (Fig. 2.1). Animal cells lack cell wall and chloroplasts, while centrioles are not found in plant cells. Further, RBCs in animals are devoid of nuclei.

2.2. CELL WALL

Plant cells are surrounded by a nonliving and rigid coat called ***cell wall***. The main functions of cell walls are to provide plant cells a definite shape, and mechanical support and strength to tissues and organs. Cells in meristematic tissues have relatively thin and elastic cell wall. The differentiation of newly produced cells into various tissues is accompanied with a thickening of cell wall due to various types of depositions. The pattern of these depositions on the inner surfaces of cell walls shows a considerable variation among different tissues. Even walls of individual cells are not uniformly thick over their entire surface; a variable frequency of pits of different types are distributed over the entire cell wall. In plants, differentiation is primarily manifested in the structure of cell wall, which shows a marked and consistent variation among various tissues. In fact, plant tissues are identified on the basis of the structure of their cell walls. The cell wall of a differentiated cell has the following three distinct parts: (1) middle lamella, (2) primary cell wall, and (3) secondary cell wall (Fig. 2.2).

2.2.1. Middle Lamella

In plants, the walls of contiguous (immediate neighbour) cells are joined by ***middle lamella***. This is composed mainly of pectin, occurring most likely in the

form of calcium (Ca^{2+}) and magnesium (Mg^{2+}) salts. Adhesion of the walls of contiguous cells is primarily dependent on the presence of Ca^{2+} and Mg^{2+} ions in the middle lamella. Pectin is readily hydrolysed by the enzyme pectinase as well as by strong acids. Cellulose is characteristically absent from middle lamella.

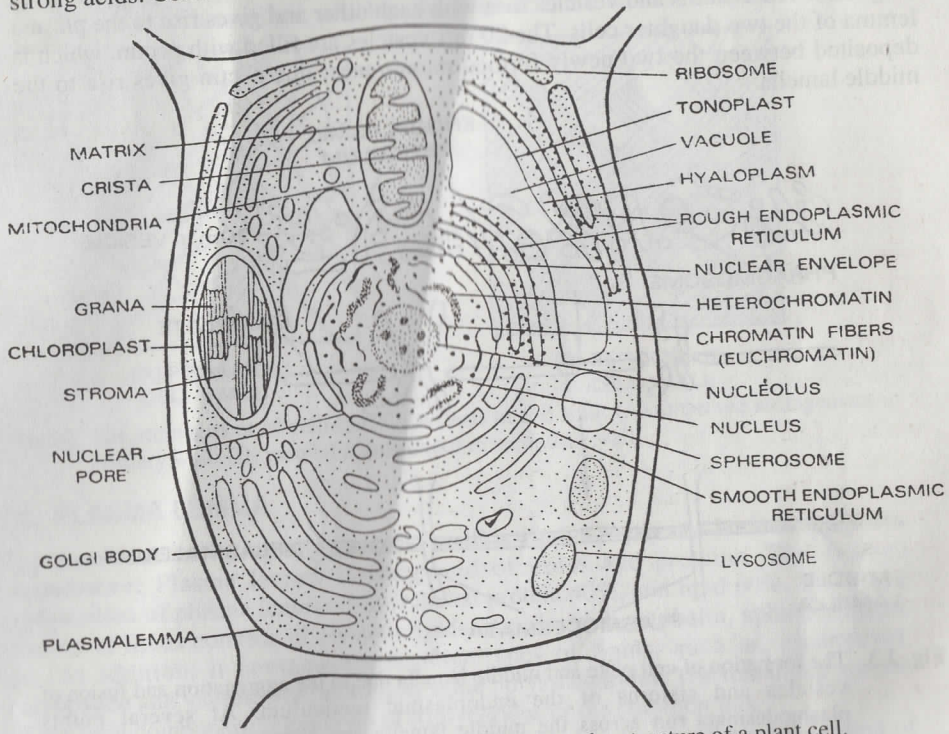


Fig. 2.1. A schematic representation of the ultrastructure of a plant cell.

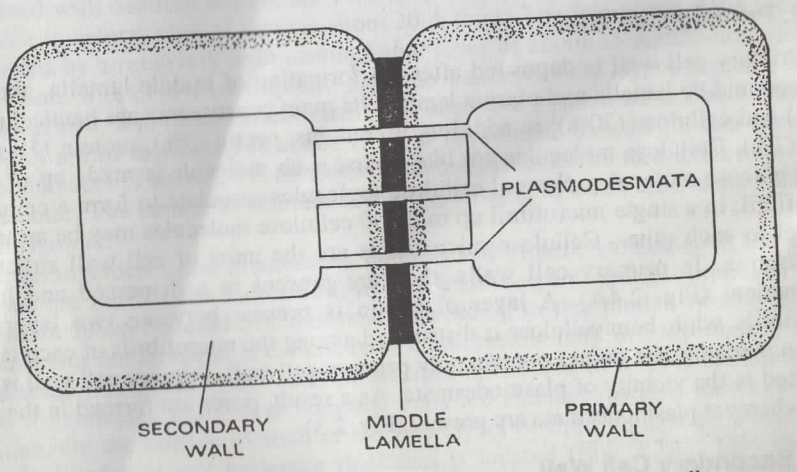


Fig. 2.2. The structure of cell wall of a mature plant cell.

Middle lamella is the first to originate following cell division. The following membranous structures accumulate at the equatorial plates of cells during telophase: (1) small cisterns of ER origin, (2) spherical vesicles of 20 mμ diameter, and (3) phragmosomes of about 250 mμ diameter. These structures are arranged in a layer (Fig. 2.3). The cisterns and vesicles fuse with each other and gives rise to the plasma lemma of the two daughter cells. The 20 mμ vesicles are filled with pectin, which is deposited between the two newly formed membranes; this pectin gives rise to the middle lamella.

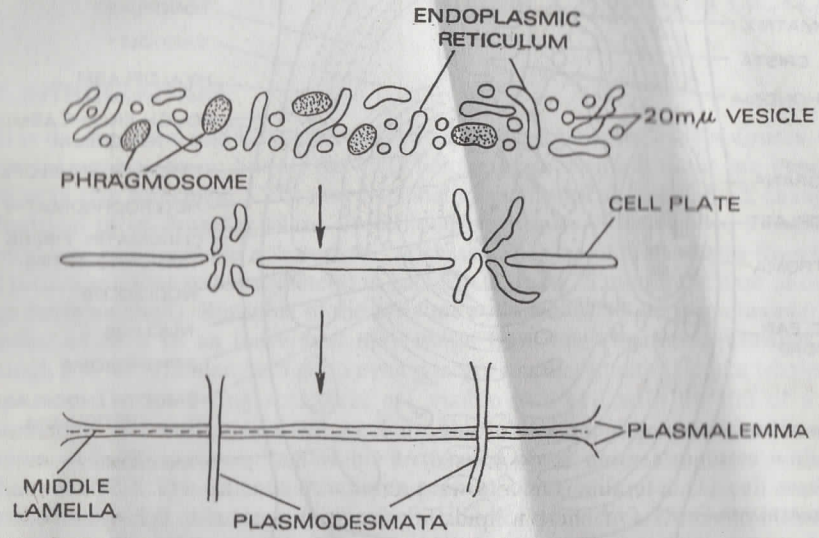


Fig. 2.3. The formation of cell plate and middle lamella due to the aggregation and fusion of vesicles and cisterns of the endoplasmic recticulum. At several points plasmodesmata run across the middle lamella and connect the cytoplasms of neighbouring cells.

2.2.2. Primary Cell Wall

Primary cell wall is deposited after the formation of middle lamella, and lies between middle lamella and plasma lemma. Its main constituents are hemicellulose (53%) and cellulose (30%); in addition, it contains, pectin (5%), protein (5%) and lipid (7%). Cellulose molecules are like fibers; each molecule is made up of upto 3,000 glucose molecules. Several cellulose molecules associate to form a cellulose microfibril; in a single microfibril up to 2,000 cellulose molecules may be arranged parallel to each other. Cellulose microfibrils are the units of cell wall structural organisation. In primary cell walls, they are present in a dispersed and loose arrangement (Fig. 2.4A). A layer of pectin is present between two layers of microfibrils, while hemicellulose is distributed among the microfibrils of each layer. Cells in meristematic tissues contain only primary cell wall. Primary cell wall is not deposited in the vicinity of plasmodesmata. As a result, pores are formed in the cell walls wherever plasmodesmata are present (Fig. 2.3).

2.2.3. Secondary Cell Wall

Secondary cell wall is the last to be deposited and lies between primary cell wall

and plasma lemma; in a cell it is the innermost layer of wall. It is composed mainly of cellulose, but in many tissues it contains lignin, suberin and, in some cases, cutin. In secondary walls, cellulose microfibrils are relatively more cleanly packed and they are arranged more or less parallel to each other (Fig. 2.4B). Several microfibrils associate to form a macrofibril, which is the structural unit of secondary cell walls.

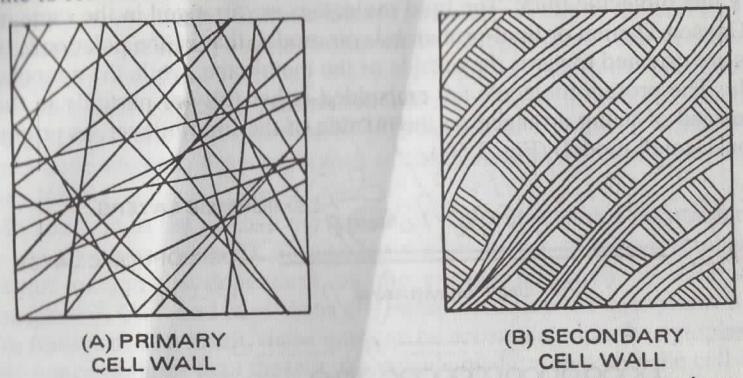


Fig. 2.4. The arrangement of cellulose microfibris in cell wall: (A) dispersed arrangement in primary wall; (B) parallel arrangement in secondary wall.

2.3. PLASMA LEMMA

The membrane enclosing cytoplasm of a cell is known as ***plasma lemma*** or ***plasma membrane***. Plasma lemma is composed of lipids and proteins. The average composition of plasma lemma is as follows: protein 40%, and lipid 60%. The major portion of lipids consists of phospholipids, *e.g.*, lecithin, cephalin, sphingomyelin, etc. In addition, it contains some other types of lipids, such as, cholesterol, cerebroside and ganglioside, as well as some polysaccharides. The remaining portion of the lipid component is made up of triglycerides or the common fats.

Three distinct layers are seen in the electron micrographs of plasma lemma of tissues fixed with osmium tetraoxide. Two of these layers are relatively dense and *osmiophilic* in nature; each of them is about 20 Å thick. The two osmiophilic layers are separated by a relatively light *osmiophobic layer* of about 35 Å thickness (Fig. 2.5 A). These three (two osmiophilic and one osmiophobic) layers together are known as a ***unit membrane***; this term was coined by Robertson. The average thickness of a unit membrane is about 75 Å. It is probable that the plasma lemma may contain pores of about 10 Å diameter, or at least it functions as if it has pores of 10 Å diameter. The structures of other membranes found in a cell are similar to that of plasma lemma.

The arrangements and orientations of lipid and protein molecules in the unit membrane are not clearly understood. In 1935, Davson and Danieli proposed the ***lipid-protein bilayer model*** of membrane. According to this model, two layers of lipid molecules are arranged in the centre of a unit membrane, each layer being only one molecule thick. The polar ends of all lipid molecules in a layer are oriented towards the outside, while their nonpolar ends are situated towards the centre of membrane. On the outside of each of the two layers of lipid molecules, a layer of protein molecules of one molecule thickness is located (Fig. 2.5A). This model agrees well with the unit membrane structure seen in electron micrographs. However,

many properties of plasma lemma and other biological membranes are difficult to explain according to this simple model.

A more flexible model, referred to as ***fluid-mosaic model***, has been proposed to account for the various properties of biological membranes. According to this model, membranes are made up of two layers of lipid molecules (lipid bilayer), each layer being only one molecule thick. The lipid molecules are arranged in the same manner as in the Davson-Danieli or lipid-protein bilayer model, that is, the polar ends of lipid molecules are oriented towards the outside of the membrane, while the nonpolar ends face inside. The protein molecule are embedded either fully or partially in the lipid bilayer (instead of being arranged on the outside of the lipid bilayer, as proposed in the Davson-Danieli model) (Fig. 2.5C).

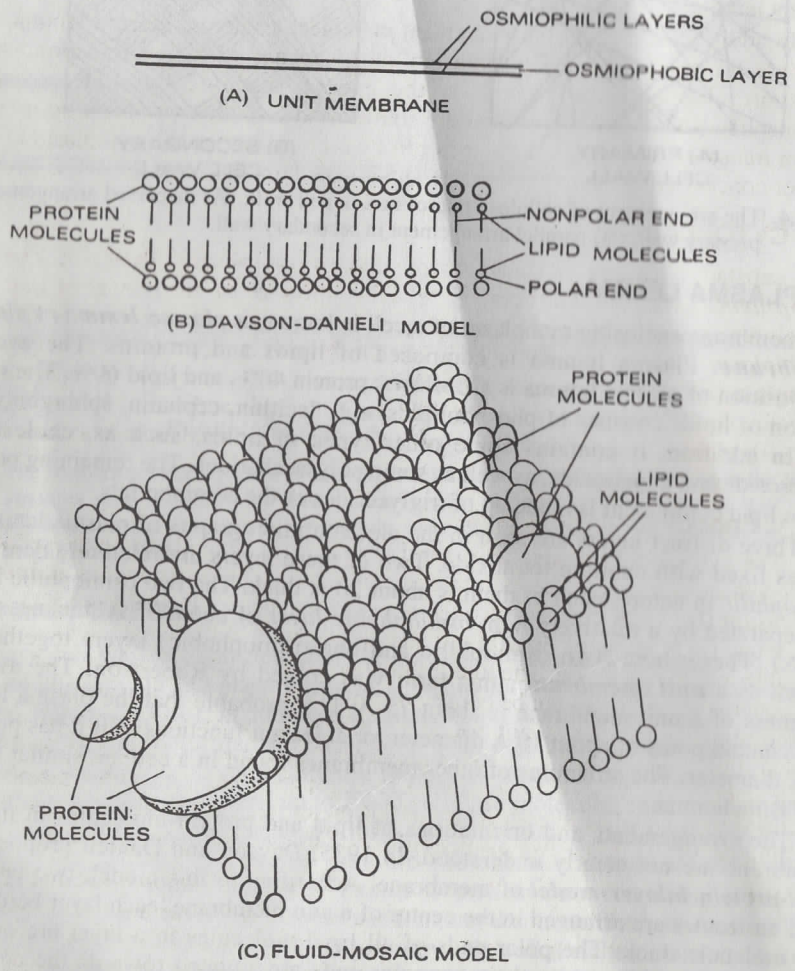


Fig. 2.5. Plasma lemma. (*A*) Unit membrane seen under electron microscope after osmium staining. (*B*) The Davson-Danielli model of the unit membrane organisation, and (*C*) the fluid-mosaic model of the unit membrane (note that in this model, protein molecules are embedded in the lipid bilayer).

The chief function of plasma lemma is to regulate the movement of various molecules into and out of the cytoplasm. Smaller molecules, like water, pass through the plasma lemma more easily than relatively larger molecules. Similarly, lipid soluble molecules pass through the plasma lemma more readily than lipid insoluble ones. Solid particles and some solutes enter the cytoplasm through phagocytosis and pinocytosis, respectively. Cellular wastes are excreted out of the cell by the process in which the sequence of events is essentially the opposite of that in pinocytosis and phagocytosis. Vesicles containing waste materials migrate to the plasma lemma, and their membranes fuse with plasma lemma at the point of contact. Eventually, the fused portion of the two membranes is disrupted as a result of which the vesicles open outside the cell; the membranes of these vesicles become integrated into the plasma lemma.

In addition to the passive movement of molecules across plasma lemma, some ions are transported across plasma lemma by means of active transport. Active transport of ions is dependent on the energy provided by ATP (adenosine triphosphate). Ca^{2+} and Mg^{2+} ions are transported into the cytoplasm by means of active transport; as a result, these ions can be accumulated in the cytoplasm in much higher concentrations than those in the outside medium in which the cell is present.

2.4. CYTOPLASM

The substance, except nucleus, surrounded by the plasma lemma is known as ***cytoplasm***. Electron micrographs reveal a number of membranous and other structures in the cytoplasm. The cytoplasm minus these structures is known as ***hyaloplasm***. The cytoplasm may contain the following structures: (1) endoplasmic reticulum (ER), (2) ribosomes, (3) golgi bodies, (4) lysososmes, (5) spherosomes, (6) vacuoles, (7) centrioles (in animals only), (8) microtubules, (9) mitochondria, and (10) plastids (in green plants only). Of these, mitochondria and plastids contain DNA (the genetic material).

2.4.1. Endoplasmic Reticulum

The cytoplasm contains an extensive network of membrane-enclosed spaces; these spaces along with the membranes enclosing them are known as ***endoplasmic reticulum*** (ER). ER consists of three types of elements: (1) vesicles of 25-500 mμ diameter, (2) tubules of 50-100 mμ diameter, and (3) 40-50 mμ thick cisterns of variable length and width. The tubules may or may not be extensively branched, and the cisterns may or may not be connected with each other. ER is generally present in differentiated cells; it is not found in RBCs, eggs, undifferentiated cells and prokaryotic cells (Table 2.1). ER is connected with plasma lemma as well as with nuclear envelop. It is believed that ER originates from the outer membrane of nuclear envelope. The ultrastructure of ER membranes is the same as that of a unit membrane, that is, it has two osmiophilic layers separated by an osmiophobic layer.

ER is grouped into the following two categories: (1) smooth ER and (2) rough ER. In the ***smooth ER*** elements, both outer and inner surfaces are regular and smooth. In those cells where little or no protein synthesis takes places, only smooth ER is found. In the ***rough ER*** elements, the outer surfaces of membranes have a rough appearance due to the attachment of ribosomes. Rough ER is mainly composed of cisterns (membrane-enclosed plate-like elements) and is found in cells actively involved in protein synthesis. Smooth and rough ER change into each other as per the needs of a cell.

The main functions of ER are as follows: (1) it provides the structural base for protein (rough ER), lipid and phospholipid (both rough and smooth ER) syntheses; (2) it provides the channel for the transport of materials synthesised in association with ER to the various parts of a cell and even outside the cell; (3) it provides a controlled passage for the export of mRNA (messenger RNA) molecules from nucleus to rough ER, and (4) several enzyme molecules, *e.g.*, glucose-6-phosphatase, ATPase etc., are embedded in the membranes of ER.

2.4.2. Ribosomes

Ribosomes are particles of about 200 Å in diameter; they are composed of RNA and protein. Generally ribosomes are attached to the outer faces of ER membranes; this converts smooth ER elements into rough ER. But some ribosomes are present in the hyaloplasm unattached to any membrane. On the basis of their sedimentation rate, ribosomes are classified into the following three groups: (1) 80 *S* size animal ribosomes, (2) 80 *S* plant ribosomes, and (3) 70 *S* eukaryotic organelle (mitochondria and plastid) and prokaryotic ribosomes.

TABLE 2.1
Structural differences between eukaryotic and prokaryotic cells

Eukaryotic cells	*Prokaryotic cells*
The nucleus is enclosed by two concentric unit membranes.	The region containing chromosome is not surrounded by a membrane.
DNA is complexed with histone and non-histone proteins to form chromatin fibers.	Generally, DNA is naked and not complexed with histones.
The amount of DNA per cell is many-fold greater than that in prokaryotes.	The amount of DNA/cell is a small fraction of that in eukaryotic cells.
Cytoplasm contains several membranous structures, *e.g.*, endoplasmic reticulum, Golgi bodies, lysosomes, etc.	Cytoplasm does not have ER and other membranous structures (except chromatopheres in photosynthetic bacteria and blue-green algae).
Cells contain mitochondria which are the sites of oxidative phosphorylation.	Mitochondria are absent; oxidative phosphorylation is associated with plasma lemma.
All green plants have typical chloroplasts, which have typical grana.	Chloroplasts are not found. Blue-green algae have lamellar photosynthetic structure but they do not have grana.
Cytoplasm has a system of microtubules, which may provide stability to the cytoplasmic structures.	Cytoplasm does not have any microtubules.
Ribosomes are of 80 *S* size; in plants, they dissociate into 60 *S* and 36 *S* subunits.	Ribosomes are 70 *S* size; they are made up of 50 *S* and 30 *S* subunits.
During cell division, spindle apparatus is organised; spindle fibers are associated with the movement of chromosomes/chromatids to the opposite poles of a cell.	Spindle fibers are absent. Chromosome movement is produced by an enlargement of the plasma lemma to which they are attached.
Chromosome ends are attached to the nuclear membrane (near annuli).	Chromosomes are attached to plasma lemma.
Most of the ribosomes are attached to endoplasmic reticulum; some ribosomes are free in the cytoplasm.	All ribosomes are free in the cytoplasm.
Nucleolus is present in the nucleus.	Nucleolus is absent.

In the conditions of low Mg^{2+} concentration, each 70 *S* prokaryotic ribosome dissociates into 50 *S* and 30 *S* subunits. The smaller subunit of ribosome is placed onto the larger subunit much like a cap-on-the-head (Fig. 2.6). These ribosomes have 37% protein and 63% RNA.

In *E. coli*, each 50 *S* larger subunit has one molecule each of a 23 *S* and a 5 *S* (about 120 nucleotides) RNA, while the 30 *S* smaller subunit has one molecule of 16 *S* RNA (about 1,600 nucleotides). The larger subunit has 34 different proteins (L_1 to L_{34}), while the smaller subunit has 21 different proteins (S_1 to S_{21}, Fig. 2.6). Detailed three dimensional models of ribosomes have been proposed but their description is beyond the scope of this book.

Ribosomes are synthesised and assembled in nucleolus (within the nucleus) from where they migrate into the cytoplasm. Ribosomes are essential for protein synthesis as mRNAs can support protein synthesis only when they are attached to ribosomes.

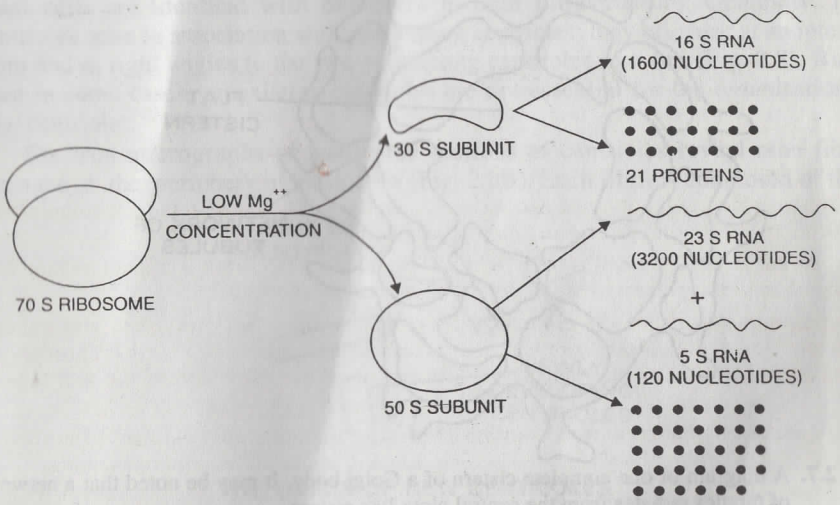


Fig. 2.6. The composition of 70 *S E. coli* ribosomes.

2.4.3. Golgi Body

It consists of 2-7 flat cisternae stacked close to each other (Fig. 2.1). A network of 300-500 Å tubules emerges from around the margins of the cisternae (Fig. 2.7). In addition, vesicles of 200-800 Å diameter are also present on the margins of the cisternae. The number of Golgi bodies in a cell depends on the synthetic activity of the cell. In a synthetically active cell, several well differentiated and developed golgi bodies are present, while synthetically inactive cells have few poorly developed Golgi bodies. Golgi bodies originate from ER elements.

It is believed that materials synthesised in association with ER (proteins, lipids, phospholipids, etc.) are transported to Golgi bodies, where they are packed into vesicles cut off from them. Thus Golgi bodies function as packaging plants of the cell.

2.4.4. Lysosomes

Lysosomes are vesicles of 400-800 mμ and contain several hydrolytic enzymes. The main enzyme present in lysosomes is acid phosphatase; other enzymes are, acid DNAase, acid RNAase, β-galactosidase, etc. The membrane surrounding a lysosome has a unit membrane organisation.

Lysosomes are of the following two types: (1) primary and (2) secondary lysosomes. ***Primary lysosomes*** are produced by Golgi bodies and contain hydrolytic ennzymes only. They fuse with food vacuoles produced through phagocytosis and pinocytosis to generate ***secondary lysosomes.*** Therefore, secondary lysosomes contain both hydrolytic enzymes as well as food materials. The food particles are ultimately digested by the hydrolytic enzymes and absorbed into the hyaloplasm, while the undigested portion of food materials remains in the secondary lysosomes; such lysosomes are known as ***residual bodies.***

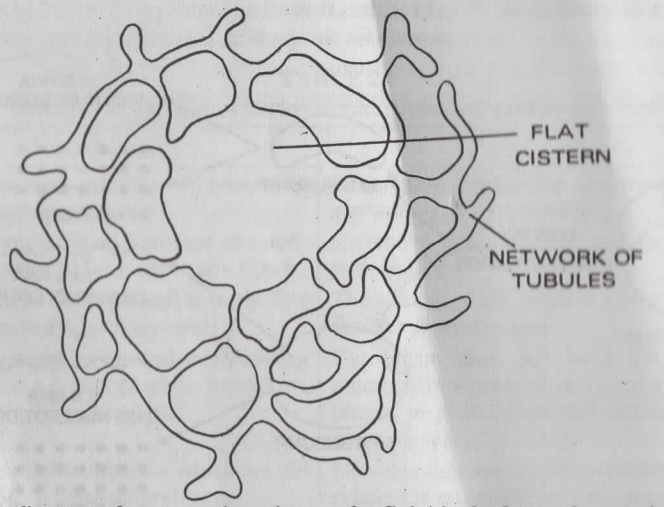


Fig. 2.7. A diagram of one complete cistern of a Golgi body. It may be noted that a network of tubules radiates from the central plate-like cistern.

The enzymes present in lysosomes are capable of digesting any living organism. When a white blood cell (WBC) ingests (by phagocytosis) a bacterium or some other organism, all the lysosomes of the WBC fuse with the food vacuole thus produced. As a result, the bacterium as well as the WBC itself are lysed. In some situations, the enzymes present in lysosomes are released into the cytoplasm, which leads to the lysis of concerned cell (***autolysis***). Thus the function of lysosomes is digestion (lysis) of food particles and microorganisms ingested by a cell and also to cause autolysis of the cell, if required.

2.4.5. Spherosome

Spherosomes are vesicles of 0.5-1 μ diameter and contain, upto 98%, lipid; they also contain some acid phosphatase. Spherosomes are found in plant cells only; they do not occur in animal and prokaryotic cells. Spherosomes are surrounded by a single unit membrane. According to some scientists, spherosomes function in lipid storage, but some others do not agree with this view.

2.4.6. Vacuole

Plant cells have one or more vacuoles of variable size. In mature and differentiated cells, the major part of cytoplasm is occupied by a large vacuole, and the cytoplasm is pushed to the periphery of cells. The material contained in the vacuoles is referred to as ***cell sap.*** Cell sap is relatively less dense than the surrounding cytoplasm. Cell sap contains sugars, salts, proteins, phenols etc. as well as some specific pigments, *e.g.*, anthocyanin. Generally, vacuoles are surrounded by a unit membrane; this membrane is referred to as ***tonoplast.*** Tonoplasts exhibit differences in permeability as compared to the plasma lemma.

2.4.7. Centriole

Centrioles are cylindrical structure of about 1200-1500 Å in diameter and of about 3,000-5,000 Å in length. They are always present in pairs, one centriole being oriented vertical to the other (Fig. 2.8). Centrioles are confined to animal cells, and are not found in plant cells. However, basal bodies found at the base of flagella of plant cells are identical with centrioles in their ultrastructure. Generally, new centrioles arise in association with preexisting centrioles; they organise at an interval from and at right angles to the two preexisting centrioles in a cell (Fig. 2.8). But at least in some cases, preexisting centrioles are not essential for the organisation of new centrioles.

Electron micrographs of transverse sections of centrioles reveal nine fibrils arranged at the periphery of a centriole (Fig. 2.8B). Each fibril is composed of three

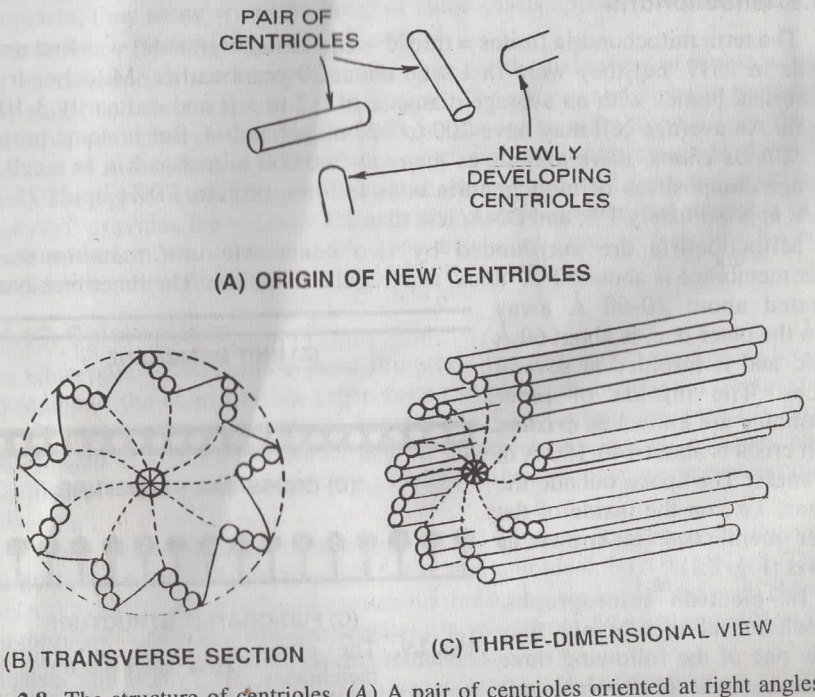


Fig. 2.8. The structure of centrioles. (*A*) A pair of centrioles oriented at right angles with each other. New centrioles are developing at right angles to the preexisting ones. (*B*) A transverse section of one centriole. (*C*) A three dimensional view of one centriole. One centriole is composed of nine groups of three microtubules each.

microfibrils arranged in a single plane so that each fibril has a flat topography. The fibrils of a centriole are oriented at an angle on the periphery. The inner microfibril of each of the nine fibrils is connected with the outer microfibril of a neighbouring fibril; at the same time, it is also connected with the microfibril located in the centre of the centriole. Thus the transverse ultrastructure of a centriole is somewhat similar to the structure of the wheel of a bullock-cart (Fig. 2.8B).

In animal cells, centrioles are involved in the organization of spindle apparatus. In fact, a pair of centrioles lies at each of the two poles in a cell from which spindle fibres radiate toward the equatorial plate. In both plant and animal cells, centrioles serve as the basal bodies of flagella.

2.4.8. Microtubules

Microtubules are tubules of 150-160 Å in diameter; their wall is about 45-70 Å in thickness. In plants, the walls of microtubules are composed of 13 fibres made up of protein molecules. It is likely that lipid molecules provide stability to the structure of microtubule walls. In centrioles and flagella, each microfibril is essentially a microtubule. Microtubules also constitute the fibres of spindle apparatus.

Microtubules are primarily found in dividing cells in the form of spindle fibers, and are responsible for chromosome movement during cell division. But they are also found it nondividing cells, and provide a degree of stability to the various structures present in the cytoplasm.

2.4.9. Mitochondria

The term mitochondria (mitos = thread + chondrion = granule) was first used by Benda in 1897, but they were first seen about 20 years earlier. Mitochondria are cylindrical bodies with an average diameter of 0.2 to 1 μ and ordinarily 3-10 μ in length. An average cell may have 200 to 800 mitochondria. But in some protozoa, *e.g., Chaos chaos,* there may be as many as 500,000 mitochondria in a cell. The average composition of mitochondria is as follows: protein, 70%; lipids 25-30%; RNA, approximately 1%; and DNA, less than 1%.

Mitochondria are surrounded by two concentric unit membranes. The outer membrane is about 60 Å thick and regular in outline. The inner membrane is situated about 20-60 Å away from the outer one, is about 60 Å thick, and is infolded at several places. The infolds of inner membrane are known as ***cristae.*** Each crista is about 140-180 Å in thickness. The space outside the cristae, *i.e.,* on the inside of the inner membrane, is known as matrix (Fig. 2.1).

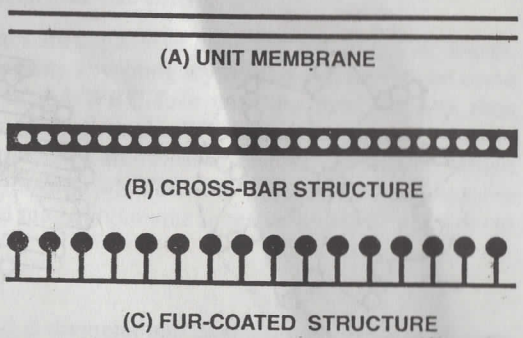


Fig. 2.9. The ultrastructure of mitochondrial membranes. (*A*) Typical unit membrane organisation. (*B*) A unit membrane with osmiophilic cross-bars. (*C*) The fur-coated structure seen after negative staining.

In electron micrographs, mitochondrial membranes may show one of the following three ultrastructures: (1) typical unit membrane organization, (2) cross-bar ultrastructure, and (3) fur-coated structure. ***The unit membrane ultrastructure*** consists of two typical osmiophilic layers separated by an osmiophilic layer (Fig. 2.9A). In the ***cross-bar ultrastructure,*** nemerous osmiophobic spheres of 40-50 Å diametre are embeded in an osmophilic layes of 75-80 Å thickness. This gives the impression as if the two osmiophobic layers of a unit membrane are connected with each other by several osmiophilic cross-bars; this structure resembles a metal or bamboo ladder (Fig. 2.9*B*). The ***fur-coated ultrastructure*** is obtained by negative staining of crista membranes with agents like phosphotungstic acid. A large number of particles of 75-100 Å in diameter are attached to the outer surface of cirsta membranes by means of 45-50 Å long stalks (Fig. 2.9*C*). It is not clear which of the three organizations is of common occurrence.

The outer membrane is considerably more permeable than the inner membrane. The inner membrane is composed of numerous identical units referred to as oxysomes. Each ***oxysome*** is capable of complete electron transport to O_2 as well as coupled oxidative phosphorylation. Clearly, the components of electron transport and coupled phosphorylation are integral parts of the inner membrane.

Mitochondria arise by fission of preexisting mitochondria. Growth and fission of mitochondria has been filmed by time-lapse photography of living cells. On the other hand, two or more smaller mitochondria may fuse together to produce a larger one.

Each mitochondrion contains several copies of ring-shaped DNA molecule. Mitochondria contain ribosomes, which are distinct from those present in the cytoplasm; they along with ribosomes of chloroplasts appear to constitute a distinct category of ribosomes (*organelle ribosomes*).

Mitochondrial matrix contains enzymes for the oxidation of amino acids and fatty acids, and the enzymes of Kreb's cycle. The enzymes of electron transport and oxidative phosphorylation are located in the inner membrane. Thus the main function of mitochondria is the oxidation of carbohydrates, amino acids and fatty acids. They produce ATP by utilizing the energy obtained from these oxidations. This ATP provides the necessary energy for various biochemical reactions of cells. Since mitochondria contain DNA, they also contribute to heredity by way of cytoplasmic inheritance.

2.4.10. Chloroplasts

Chloroplasts contain the most important pigment on earth, *i.e.,* chlorophyll. They impart the characteristic green colour to plants and carry out photosynthesis, the ultimate source of all organic compounds. Chloroplasts are typically biconvex lens-shaped of about 5 μ diameter and 3μ thickness. An average cell may have 20-40 chloroplasts, but some algae, *e.g., Chlamydomonas,* have a singe chloroplast per cell.

The average chemical composition of chloroplasts may be as follows: protein 50-59%, lipid 21-34%, chlorophyll 5-8%, carotenoids 0.7-1.1%, RNA 1-7.5% and DNA 0.2-1%. Chlorophyll and carotenoid molecules are associated with chloroplast membranes (grana and stroma lamellae); some (about 5%) chlorophyll molecules appear to be complexed with protein molecules.

Chloroplasts are enclosed by two concentric membranes, each being about 50 Å thick. The Membrane-free space enclosed by these two membranes is referred to as ***stroma.*** Within the stroma are embedded about 40-80 grana of 0.2-0.8 μ diameter

(Fig. 2.1). Each granum contains 5-25 flat cisternae (membrane-enclosed spaces) stacked on top of each other. Most cisternae are confined to a single granum; such cisternae are called ***grana loculi***. But some cisternae interconnect two or more grana; they are referred to as ***fret channels***. The portion of cisternae that pass through storma (between two grana) constitutes the ***stroma lamellae***. Grana and storma lamellae appear to be composed of several thousand identical units, called ***quantasomes***. A group of 3-8 quantasomes is capable of electron transport and photophosphorylation. But the enzymes of the dark reaction of photosynthesis are located in the storma of chloroplasts.

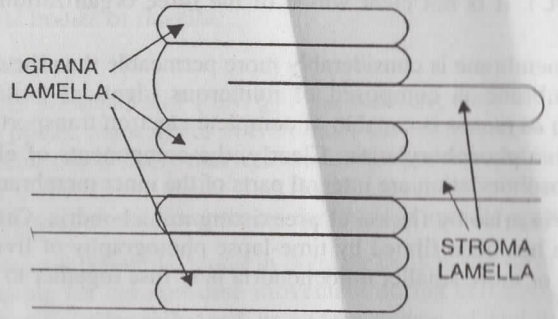


Fig. 2.10. The arrangement of grana and stroma lamellae in a granum. Note that the membranes within the granum are twice as thick as those in the stroma or those at the peripheri of the granum.

Several copies of a ring-shaped DNA molecule are present in the stroma of each chloroplast. Chloroplasts contain ribosomes and tRNA, and carry out protein synthesis. Chloroplasts differentiate from small vesicles enclosed by two concentric membranes; these vesicles are referred to as ***proplastids***. Proplastids divide by fission to produce more proplastids. Chloroplasts may also arise by fission of preexisting, differentiated chloroplasts.

In addition to chloroplasts, plant cells contain two more types of plastids called (1) leucoplasts and (2) chromoplasts. ***Leucoplasts*** are colourless plastids and function in the storage of starch, proteins and fats. ***Chromoplasts,*** on the other hand, contain pigments other than chlorophyll, *e.g.,* phucoxanthin, phycocyanin etc. The different types of plastids are obtained through differentiation from proplastids. In addition, plastids of one type readily transform into the other two types as per the needs of cells (Fig. 2.11).

Chloroplast differentiation from proplastids is a light-dependent phenomenon. In the presence of light, the inner membranes of proplastids invaginate, and a large number of small vesicles bud-off from these invaginations into the stroma. Later, these vesicles become oriented in regular arrangements; the vesicles lying in one plane fuse together to generate grana and storma lamellae. The sequence of events in the presence of high and low light intensities are markedly different, but the final outcome of the two sequences is quite comparable.

The main function of chloroplasts is photosynthesis. Light energy liberates electrons from chlorophyll molecules; these electrons pass through the electron transport chain located in chloroplast membranes and the energy derived from this transport is used to produce ATP (***photophoshorylation***). This ATP is utilized for the fixation and reduction of CO_2; this is accomplished by the soluble enzymes

located in chloroplast stroma. In addition, chloroplasts contribute to heredity in the form of cytoplasmic inheritance since they contain DNA.

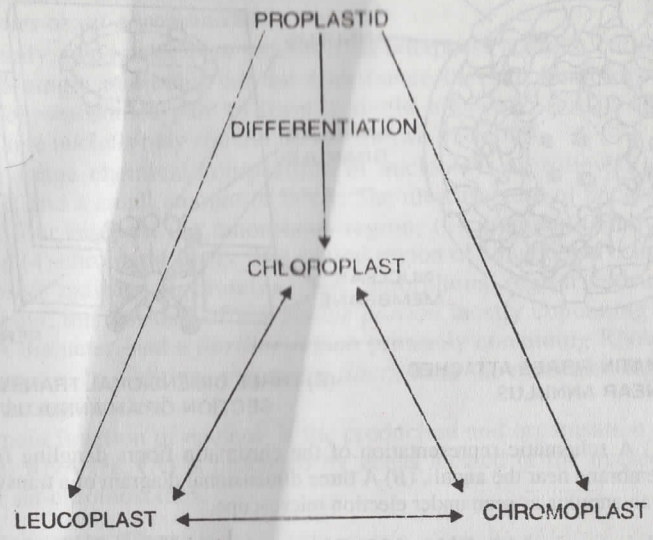


Fig. 2.11. The interrelationship among proplastids, chloroplasts, chromoplasts and leucoplasts.

2.5. NUCLEUS

Nuclei take up relatively deep stain with basic dyes, and are generally spherical in shape. But in differentiated tissues, they may exhibit considerable variation in shape. The size of nuclei varies with the physiological state and the degree of differentiation of the cells. In physiologically active and undifferentiated cells, nuclei are relativeley larger in size, while in physiologically inactive and differentiated cells they are relatively smaller. Nuclear size is also affected by the chromosome number (or the amount of DNA) of a species, as well as by the ploidy level of a cell or individual. In general, a cell contains a single nucleus, but in some tissues, *e.g.,* early stages of endosperm development, and in some organisms, *e.g.,* some protozoa, some fungi etc., two or more nuclei are present within a single cell. In contrast, some specialized cells, *e.g.,* RBCs and companion cells of phloem, do not contain any nucleus.

Nucleus is the store-house of almost all the genetic information needed for the functioning of a cell/organism. It produces ribosomes, tRNA (transfer RNA), mRNA (messenger RNA) and certain other RNAs, which produce the various structural and enzymatic proteins. In essence, nucleus governs the development of almost all the traits of an organism by providing the information necessary for the synthesis of various structural enzymatic and regulator proteins. The significance of nuclei in development was first demonstrated by a German scientist, Hammerling in 1934, through transplantation studies in the unicellular alga *Acetabularia*. He clearly demonstrated hat the development of characters in this alga was governed by the nucleus and not by the cytoplasm.

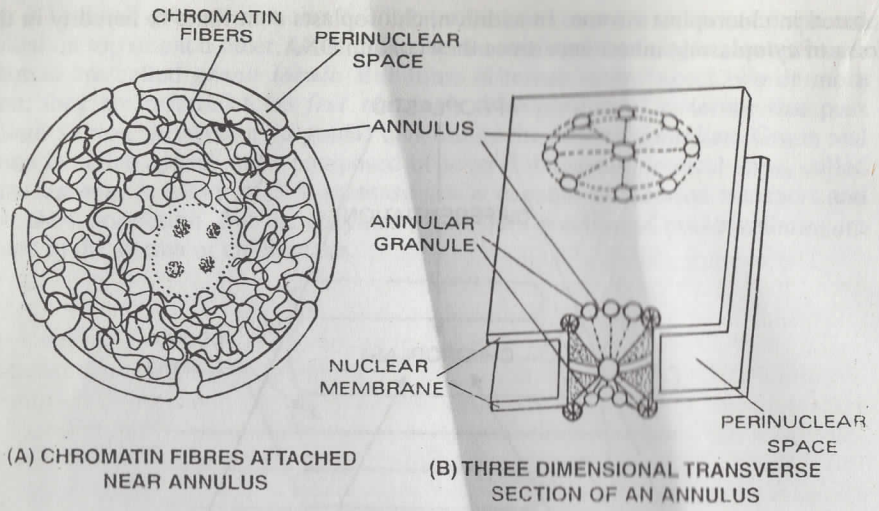


Fig. 2.12. (*A*) A schematic representation of the chromatin fibers dangling from nuclear membrane near the annuli. (*B*) A three dimensional diagram of a transverse section of an annulus as seen under election microscope.

Pea nuclei contain 14% DNA, 12% RNA and 74% protein on dry weight basis. In electron micrographs, the following three distinct types of structures are discernible in interphase nuclei: (1) nuclear envelope, (2) nucleolus, and (3) chromatin fibers (Fig. 2.1 and 2.12A)

2.5.1. Nuclear Envelope

Nucleus is enclosed by two concentric membranes, each being 70-80 Å in thickness. There is a space of about 200-300 Å between the two membranes; this space is known as the ***perinuclear space***. The two membranes, together with the perinuclear space, are known as ***nuclear envelope***. In the nuclear envelope, ***nuclear pores*** of 200-400 Å in diameter are arranged in a hexagonal pattern at an interval of 800 Å from each other. In fact, ***nuclear pores*** or ***annuli*** are not real pores in that the membrane is continuous throughout the area of an annulus. The two membranes of nuclear envelope are fused with each other at the peripheri of each annulus (Fig. 2.12*A*). At the peripheri of an annulus, eight annular granules are embedded in each of the two nuclear membranes. In the centre of the annulus, another annular granule is located. The annuli provide the main channel for transport of materials from and into the nucleus.

2.5.2. Nucleolus

Interphase nuclei contain relatively dense spherical bodies called ***nucleolus***. Generally, nucleoli disappear during cell division, in the late prophase of mitosis and meiosis, and the nucleolar material is distributed onto the chromosomes. At the end of cell division, the nucleolar materials dissociate from chromosomes and fuse together and become organised into a new nucleolus.

Nucleolus is ordinarily attached to a specific region (***nucleolus organiser region***) of a specific chromosome (***nucleolus organiser chromosome***). At mitotic metaphase, the nucleolus organiser region generally appears as an additional

constriction or even as a gap; this region therefore is known as ***secondary constriction***. The part of chromosome beyond the secondary constriction appears as a satellite; such chromosomes, therefore, are often referred to as ***satellite chromosomes*** or ***sat-chromosomes***.

Ordinarily, two nucleoli are present in an interphase nucleus, but during meiotic prophase a single nucleolus can be seen (since the two homologous nucleolus organiser chromosomes pair to form a single bivalent). But in certain special situations, one nucleus may contain more than two nucleoli.

The average chemical composition of nucleoli is as follows: protein, 70%; RNA, 20%; and a small amount of DNA. The ultrastructure of nucleoli reveals the following four regions: (1) amorphous region, (2) granular region, (3) fibrillar region, and (4) chromatin fibers. The central region of a nucleolus is amorphous and is composed mainly of proteins. The amorphous region is surrounded by nucleolonema; this consists of a *granular portion* mostly containing ribosomes of 150-200 Å diameter, and a *fibrillar region* primarily containing RNA fibers of 50-150 Å diameter. In addition, *chromatin fibers* from the nucleolar organiser region also pass through nucleoli.

The main function of nucleoli is the production and organisation of ribosomes. Ribosomal RNA (rRNA) is produced by genes located in the nucleolar organiser regions of sat-chromosomes.

2.5.3. Chromatin Fibers

In transverse sections, transversely and obliquely-cut ends of chromatin fibers are seen throughout the interphase nucleus, except in the major part of nucleolus. The average diameter of chromatin fibers is about 300 Å. In interphase nuclei, chromatin fibers are relatively much more condensed in heterochromatin than those in euchromatin. As a result, heterochromatic regions appear much more dense than euchromatic regions (Fig. 2.1).

During interphase, one end (telomore) of each chromosome is attached at the peripheri of an annulus. The rest of the chromosome dangles free in the nucleus in a highly, but loosely, folded state (Fig. 2.12B). Thus interphase nuclei are filled with a mass of folded chromatin fibers. Chromatin fibers contain about 40% DNA, 55% protein and 4-5% RNA. They are the basic units of chromosome structure. They are also the fundamental basis of inheritance since they contain the genetic material, *viz.*, DNA. A more detailed discussion of chromatin fibers follows in the next chapter (Chapter 3).

QUESTIONS

1. List the various structures seen in an electron micrograph of a generalised plant cell. Describe the ultrastructure of the following with the help of suitable diagrams:
   (*i*) cell wall,
   (*ii*) plasma lemma, and
   (*iii*) endoplasmic reticulum.
2. Describe the ultrastructure and function of the following structures present in the cytoplasm of a cell:
   (*i*) ribosomes,
   (*ii*) lysosomes,
   (*iii*) microtubules, and
   (*iv*) vacuoles.

3. Describe the ultrastructure, origin and function of plastids (or mitochondria) with the help of suitable diagrams.
4. Describe the ultrastructure of an interphase nucleus with the help of suitable diagrams. Discuss the functions of a nucleus.
5. Write short notes on the following:
   (*i*) nucleolus, (*ii*) nucleus,
   (*iii*) chloroplast, (*iv*) chromatin fibers,
   (*v*) mitochondria, (*vi*) Golgi bodies,
   (*vii*) endoplasmic reticulum, (*viii*) ribosomes,
   (*ix*) annulus.

3

Chromosome Structure and Function

3.1. INTRODUCTION

Chromosomes are the rod-shaped, dark-stained bodies as seen during metaphase stage of mitosis when the cells are stained with a suitable basic dye and viewed under a light microscope. Chromosomes were first described by Strausberger in 1875. The term 'chromosome', however, was first used by Waldeyer in 1888. *They were given the name chromosome* (chroma = colour + soma = body) *due to their marked affinity for basic dyses.* As a consequence of this, they are stained rather deeply, while the remaining cytoplasm remains relatively unstained.

Chromosome
- Chroma (= colour) + soma (= body).
- First described by Strausberger (1875)
- Term 'chromosome' coined by Waldeyer (1888)
- Stain deeply with basic dyes.
- Appear rod-shaped during mitotic metaphase.
- Morphology changes with the stage of cell division.
- Chromosome integrity during interphase was shown by Boveri (1890).

Chromosomes are clearly visible as distinct bodies during the stages of cell division only. Their number can be counted with relative ease only during mitotic metaphase. The interphase nucleus, on the other hand, does not exhibit any structure resembling chromosomes under the light microscope. Studies by Boveri in 1890 and subsequently by other workers clearly established that the integrity of metaphase chromosomes is maintained through the interphase, *i.e.,* they do not disintegrate during interphase.

Chromosomes are composed of thin chromatin threads called ***chromatin fibers***. These fibers undergo folding, coiling and supercoiling during prophase so that the chromosomes become progressively thicker and smaller. Therefore, chromosomes become readily observable under light microscope. At the end of cell division, on the other hand, the fibers uncoil and extend as fine chromatin threads, which are not visible under light microscope. But under electron microscope, the interphase nuclei appear to be filled with about 300 Å thick chromatin fibers. These chromatin fibers fold in various ways to yield the typical metaphase chromosomes (Fig. 3.1, 3.4).

3.2. CHROMOSOME NUMBER

Each species has a definite and, generally, a constant somatic and gametic

chromosome number. ***Somatic chromosome number*** is the number of chromosomes found in somatic, more specifically meristematic, tissues of a species and is represented by $2n$. Ordinarily, somatic cells contain two copies of each chromosome. The two copies of a chromosome are ordinarily identical in morphology, gene content and gene order; they are known as ***homologous chromosomes***.

Chromosome Number

- Constant for each species.
- Somatic number, $2n$; two homologues of each chromosome.
- Gametic number, n; exactly one-half of somatic number.
- Lowest somatic number ($2n$).
  - (*i*) *Parascaris equorum*, (animal) = 2;
  - (*ii*) *Haplopappus gracilis* (plant) = 4.
- Range of $2n$ = 2 to 1,600 (some protozoa).
- Higher in dicots than in monocot

Gametic chromosome number is precisely one half of the somatic number and is represented by n. It denotes the number of chromosomes found in the gametes of a species. It is important to note that gametes of a species ordinarily contain one (*and only one*) copy of each of the different chromosomes of that species. Since a zygote is produced by fusion of one male and one female gamete, the zygote contains $2n$ chromosomes (somatic number). Thus *in any individual, one homologue of each chromosome comes from its male parent, while the other homologue is contributed by its female parent.*

Gametic and somatic chromosome numbers show a large variation among different species (Table 3.1). On one extreme, *Ascaris megalocephala univalens* shows only two chromosomes in its generative tissues, while the other extreme is represented by protozoa having 300 or more chromosomes; some protozoa have as many as 1,600 chromosomes. However, most organisms have somatic chromosome numbers between 12 and 50 (n ranging between 6 and 25). *Haplopappus gracilis*, a member of the family compositae, has four chromosomes in its somatic tissues, which is the lowest chromosome number known in plants. *The most commonly found gametic number* (*n*) *ranges from* 3-8 *in fungi, from* 8-16 *in angiosperms* (*the most common number being* 12), *and from* 16 to 30 *in primates.*

3.3. CHROMOSOME SIZE

In contrast to other cell organelles, *the size of chromosomes shows a remarkable variation depending upon the stage of cell division. Chromosomes are the longest and the thinnest during interphase, so that they are not even visible under light microscope. With the onset of prophase, there is a progressive decrease in their length accompanied with an increase in thickness* (*diameter*). *Chromosomes are the smallest during anaphase. But chromosomes are the most easily observed and studied during mitotic metaphase when they are very thick, quite short and well spread in the cell. Therefore, chromosome measurements are generally taken during mitotic metaphase.*

TABLE 3.1
Somatic and gametic chromosome numbers of some common plants and animals

Scientific name	*Common name*	*Chromosome number*	
		Somatic	*Gametic*
	Plants		
Allium cepa	onion	16	8
Aspergillus nidulans	mold	16	8
Cajanus cajan	pigeonpea	22	11
Cicer arietinum	chickpea	16	8
Crepis capillaris	—	6	3
Daucus carota	carrot	20	10
Haplopappus gracilis	—	4	2
Hordeum vulgare	barley	14	7
Lycopersicon esculentum	tomato	24	12
Neurospora crassa	bread mold	14	7
Nicotiana tabacum	tobacco	48	24
Oryza sativa	rice	24	2
Penniseturm americanum	pearl millet	14	7
Phaseolus vulgaris	rajma, kidney bean	22	11
Piper nigrum	pepper	128	64
Pisum sativum	pea	14	7
Secale cereale	rye	14	7
Solanum tuberosum	potato	48	24
Sorghum bicolor	jowar	20	10
Triticum monococum	diploid wheat	14	7
Triticum durum	durum wheat	28	14
Triticum aestivum	bread wheat	42	21
Vicia faba	faba bean	12	6
Zea mays	maize	20	10
	Animals		
Apis mellifera	honey bee	32	16
Ascaris megalocephala univalens (renamed as *Parascaris equorum*)	round worm horse nematode	2	1
Bombyx mori	silkworm	56	28
Bos taurus	cattle (cow)	60	30
Canis familiaris	dog	78	39
Columbia livia	pigeon	80	40
Drosophila melanogaster	fruit fly	8	4
Equus asinus	donkey	62	31
Equus caballus	horse	64	32
Felis domesticus	cat	38	19
Gallus gallus domesticus	chicken	78	39
Homo sapiens sapiens	man	46	23
Musca domestica	house fly	12	6
Rattus norvegicus	rat	42	21

Chromosome Size
- Changes drastically with the stage of cell division.
- Measured during mitotic metaphase.
- Range : 0·5 μ to 32 μ.
- Longest in *Trillium*.
- Longer in plants than in animals.
- Longer in monocots than in dicots.

The size of mitotic metaphase chromosomes of animal and plant species generally varies between 0.5 μ and 32 μ in length, and between 0.2 μ *and* 3.0 μ *in diameter.* The longest metaphase chromosomes are found in *Trillium*: its longest chromosome is 32 μ. The giant chromosomes found in diptera (Section 3.12.2) are permanently in premetaphase stage, and they may be as long as 300 μ and upto 10 μ in a diameter. Therefore, giant chromosomes are easily visible even in the interphase nucleus.

In general, plants have longer chromosomes than animals, and species having lower chromosome numbers have longer chromosomes than those having higher chromosome numbers. Among plants, dicots, in general, have a higher number of chromosomes than do monocots; they also tend to have shorter chromosomes than the latter.

3.4. CHROMOSOME MORPHOLOGY

Chromosome appearance (morphology) changes with the stage of cell division. Mitotic metaphase chromosomes are the most suitable for studies on chromosome morphology (Fig. 3.1). In such chromosomes, the following structural features (except chromomere) are seen under the light microscope: (1) chromatid, (2) centromere, (3) telomere, (4) secondary constriction and satellite, and (5) chromomere.

Chromatid
- At telophase, each chromosome has one chromatid.
- At mitotic metaphase, longitudinal half of a chromosome.
- Has one long DNA molecule.
- Replicates semiconservatively during *S* phase of interphase.
- Unit of chromosome structure.

3.4.1. Chromatid

Each metaphase chromosome appears to be longitudinally divided into two identical parts each of which is known as ***chromatid***. The two chromatids of a chromosome appear to be 'joined' of 'fused' (in fact, they are only 'held together' rather closely) together at a point called ***centromere***. The two chromatids of a chromosome separate from each other during mitotic anaphase (and during anaphase II of meiosis) and move to the opposite poles. As a consequence, each chromosome is represented by a single chromatid during telophase.

The DNA of each telophase chromosome (composed of a single chromatid) replicates during the synthesis (*S*) phase of interphase. This produces an identical copy of the chromatid so that during prophase and metaphase each chromosome is

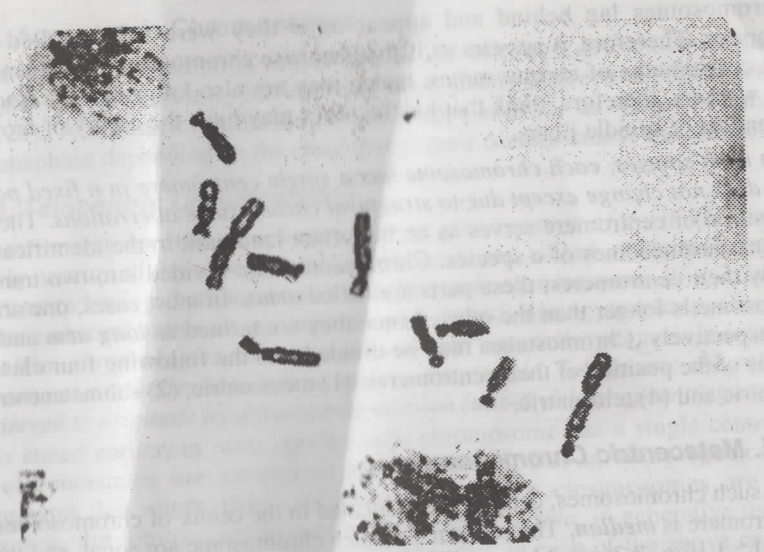


Fig. 3.1. Chromosomes of *Vicia hajastana* as seen under a light microscope during mitotic metaphase.

made up of two chromatids. Since the two chromatids making up a prophase chromosome are produced through replication of a single chromatid, they are referred to as ***sister chromatids***. In contrast, the chromatids of homologous chromosomes are referred to as ***nonsister chromatids*** (why ?).

It is almost universally accepted that ***chromatid is the structural and functional unit of chromosomes and that it is not further subdivisible into smaller subunits without adversely affecting its structural integrity and functional capability.***

Centromere
- Site of spindle fiber attachment; responsible for chromosome movement.
- Visible as constriction/gap.
- Syn., primary constriction, kinetochore.
- *Location* : median, submedian, subterminal, terminal.
- Contains highly repetitive sat - DNA; CEN sequence identified in yeast.

3.4.2. Centromere

The region where two sister chromatids of a chromosome appear to be 'joined' or 'held together' during mitotic metaphase is known as ***centromere***. Under light microscope, centromere generally appears as a constriction (a narrowed region) in the chromosome (Fig. 3.1). Therefore, it is also termed as ***primary constriction***. But sometimes, centromeric regions do not take up any stain and they appear as gaps in the chromosome.

During cell division, *spindle fibers* (*more particularly, chromosomal fibers*) *attach to centromeres. As a result, centromeres are the first parts of chromosomes to be seen moving towards the opposite poles during anaphase.* The remaining regions

of chromosomes lag behind and appear as if they were being pulled by the centromere. Therefore, it appears as if the anaphase chromosome movement is due to the centromeres of chromosomes, hence they are also known as ***kinetochores***. Some workers, therefore, think that kinetochores may have the ability of movement independent of spindle fibers.

In most species, each chromosome has a single centromere in a fixed position, which does not change except due to structural chromosome aberrations. Therefore, the position of centromere serves as an important landmark in the identification of different chromosomes of a species. Chromosomes are divided into two transverse parts by their centromeres; these parts are called ***arms***. In most cases, one arm of a chromosome is longer than the other, hence they are termed as ***long arm*** and ***short arm***, respectively. Chromosomes may be divided into the following four classes on the basis of the position of their centromeres: (1) metacentric, (2) submetacentric, (3) acrocentric and (4) telocentric.

3.4.2.1. Metacentric Chromosomes

In such chromosomes, centromere is located in the centre of chromosomes, *i.e.*, the centromere is ***median***. The two arms of such chromosome are equal, and the arm ratio is 1 : 1 [Fig. 3.2(*a*)]. Metacentric chromosomes appear as 'V'-shaped during anaphase

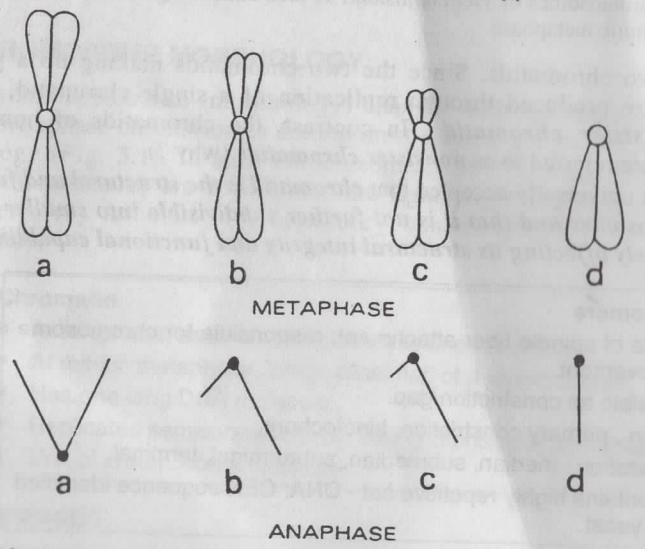


Fig. 3.2. Metaphase and anaphase configurations of the four classes of chromosomes: (*a*) metacentric, (*b*) submetacentric, (*c*) acrocentric, and (*d*) telocentric chromosomes.

3.4.2.2. Submetacentric Chromosomes

When the centromere is located on one side of the central point of a chromosome, *i.e.*, the centromere is *submedian*, such chromosomes are known as ***submetacentric*** (Fig. 3.2*b*). Such chromosomes appear either as 'V' or 'J' during anaphase, depending on how close or far-removed their centrosmeres are from their central points.

3.4.2.3. Acrocentric Chromosomes

Centromeres located close to one end of chromosomes are known as ***subterminal,*** and the chromosomes having them are called ***subtelocentric*** or ***acrocentric*** (Fig. 3.2*c*). Acrocentric chromosomes may appear either as 'j' or 'rod-shaped' during anaphase depending on the closeness of their centromeres to the telomere.

3.4.2.4. Telocentric Chromosomes

Centromeres located at one end of the chromosome (*i.e.*, in the position normally occupied by one of the telomeres) are called ***terminal***, and the chromosomes having them are known as ***telocentric chromosomes*** (Fig. 3.2*d*). The telocentric chromosomes always appear 'rod-shaped' during anaphase. *Generally, telocentric chromosomes are unstable. Most naturally occurring telocentic chromosomes are believed to be, in reality, acrocentric. Telocentric chromosomes are believed to originate by a transverse division* (***mis-division***) *of centromere.*

As stated earlier, in most species each chromosome has a single centromere; such chromosomes are termed as ***monocentric.*** But in some species, each chromosome has more than one centromere; such chromosomes are called ***polycentric***. Polycentric chromosomes are found in *Luzula,* in generative tissues of *Ascaris megalocephala univalens* and in *Thyanta*. In each of the above cases, the centromeric property is confined to one or more definite points of the chromosomes so that such centromeres are referred to as ***localized***. However, in many insects, *e.g.*, most homopteran and hemipteran insects, the centromeric activity is nonlocalized and spread over the entire chromosome length; such centromeres are known as ***diffuse centromeres***.

The structure of centromere shows considerable variation under light microscope. Lima-de-Faria proposed three models of centromere structure on the basis of extensive studies of chromosomes during meiosis. According to these models, centromeres are made up of either thin fibrils almost invisible under the light microscope (Fig. 3.3*c*) or of four or five kinetochore granules in addition to the thin fibrils (Fig. 3.3*a, b*).

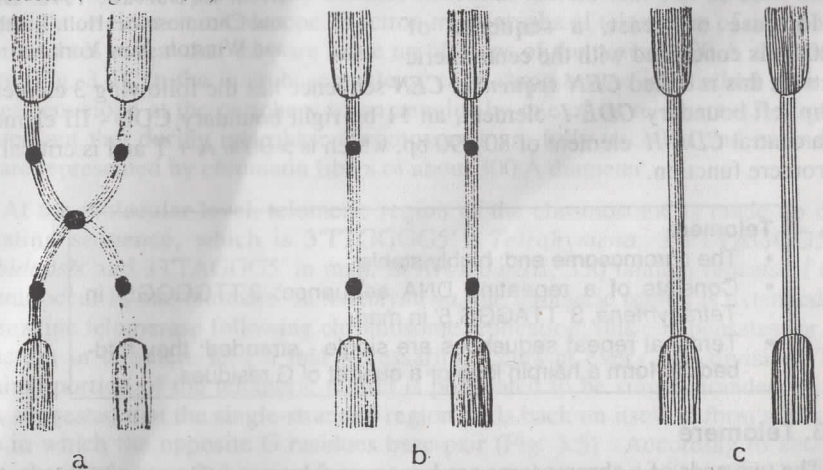


Fig. 3.3. Models of the structure of centromere as proposed by Lima-de-Faria (*a*) Kinetochore having five kinetochore granules, (*b*) a centromere with four kinetochore granules and (*c*) a centromere without any kinetochore granules.

Under scanning electron microscope, centromere appears as a relatively much thinner region than the rest of the chromosomes (Fig. 3.4). Further, the two sister chromatids are only 'held together' in the centromeric region due to adhesion; they are often separated in the centromeric region as well. Centromeres appear to be composed of the same 300 Å chromatin fibers, which are the unit of chromosome organisation, and no kinetochore granules *are visible.* .

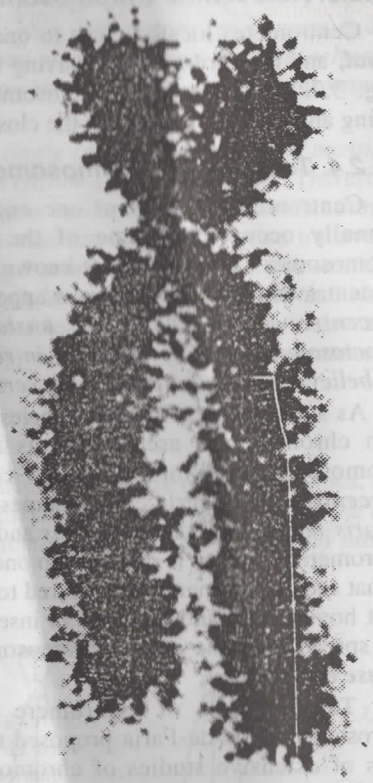


Fig. 3.4. Scanning electron micrograph of a human metaphase chromosome (from E. DuPraw, 1970. DNA and Chromosome. Holt, Rinehart and Winston, New York.).

In thin sections of chromosomes, the ultrastructure of centromere reveals relatively more densely packed 300 Å chromatin fibers and two spindle attachment granules one in each of the two sister chromatids. The centromere granules are dense bodies to which spindle fibers attach during cell division.

Centromeres contain highly repetitive DNA called ***satellite-DNA*** or ***sat-DNA***. This class of DNA usually forms one or more minor bands distinct from the rest of the chromosomal DNA under cesium chloride density gradient ultracentrifugation. In most cases, centromeric region is almost the last segment of chromosomes to replicate during late *S* phase.

In case of yeast, a sequence of ~ 120 bp is concerned with the centromeric function; this is called ***CEN sequence.*** *CEN* sequence has the following 3 elements: a 9 bp left boundary *CDE-I* element, an 11 bp right boundary CDE - III element, and a central *CDE-II* element of 80 - 90 bp, which is > 90% A + T and is critical for centromere function.

Telomere

- The chromosome end; highly stable
- Consists of a repeating DNA sequence: 3'TTGGGG5' in *Tetrahymena,* 3' TTAGGG 5' in man.
- Terminal repeat sequences are single - stranded; they fold-back to form a hairpin loop or a quartet of G residues.

3.4.3. Telomere

The two ends of a chromosome are known as ***telomeres.*** They are highly stable, and telomeres of different chromosomes do not fuse or unite. But when telomeres are damaged or removed due to chromosome breakage, the damaged chromosome

ends are highly unstable; such ends readily fuse or unite with broken ends of other chromosomes. It is generally accepted that the structural integrity and individuality of chromosomes is maintained due to telomeres.

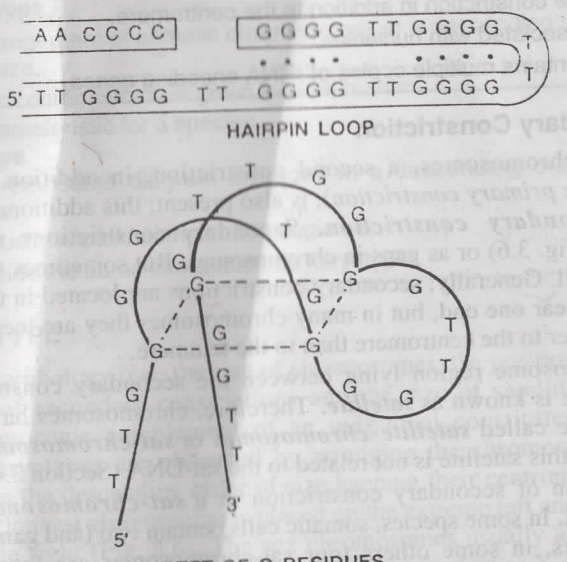


Fig. 3.5. The single-stranded region (G-T rich strand) is folded to form a hairpin, which involves interaction between G base pairs. Alternatively, the second G from 4 repeating units may form a quartet of G residues.

At the level of chromosome morphology, telomeres are only conceptual entities in that they do not represent any distinct structural feature that can be seen under either light or electron microscope. Electron micrographs of telometers of metaphase chromosomes reveal that they are made up of loops of the typical 300 Å chromatin fibers (Fig. 3.4). In the interphase nucleus, each chromosome is attached with the nuclear envelope at the periphery of an annulus by one of its telomeres. It may be pointed out that during interphase chromosomes are fully uncoiled and extended, and are represented by chromatin fibers of about 300 Å diameter.

At the molecular level, telomeric region of the chromosome is made up of a repeating sequence, which is 3'TTGGGG5' is *Tetrahymena,* 3'TTTAGGG5'in *Arabidopsis* and 3TTAGGG5' in man. In *Arabidopsis*, 350 tandem repeats of this subunit occur in one telomere. In *tetrahymena,* the telomeric region is extended by the enzyme telomerase following chromosome replication; this compensates for the reduction in telomere length that necessarily occurs at every cell division. The terminal portion of the telomeric region is postulated to be single-stranded. It has been suggested that the single-stranded region folds back on itself to form a hairpin loop in which the opposite G residues base-pair (Fig. 3.5) . According to another view, the single-stranded region forms a more complex organization (Fig. 3.5). In any case, the end of telomeres consist of a folded DNA strand; as a result, these ends

are stable and resistant to exonuclease digestion. Exonuclease enzymes digest DNA/RNA from their ends by removing one nucleotide at a time.

Secondary Constriction
- The constriction in addition to the centromere.
- Associated with nucleolus.
- Contains multiple copies of rRNA encoding genes.

3.4.4. Secondary Constriction

In some chromosomes, a second constriction, in addition to that due to centromere (the *primary constriction*), is also present; this additional constriction is known as ***secondary constriction.*** Secondary constrictions may appear as constrictions (Fig. 3.6) or as gaps in chromosomes. But sometimes they may not be observable at all. Generally, secondary constrictions are located in the short arm of chromosomes near one end, but in many chromosomes they are located in the long arm and/or nearer to the centromere than to the telomere.

The chromosome region lying between the secondary constriction and the nearest telomere is known as ***satellite.*** Therefore, chromosomes having secondary constrictions are called ***satellite chromosomes*** or ***sat-chromosomes.*** [It may be pointed out that this satellite is not related to the sat-DNA (Section 3.4.2)]

The position of secondary constriction in a ***sat-chromosome*** is fixed and remains constant. In some species, somatic cells contain two (and gametes have one) sat-chromosomes, in some others four sat-chromosomes are found (*e.g., Vicia hajastana*; Fig. 3.6), while in some others six or more such chromosomes are present (*e.g.,* human somatic cells have 10 sat-chromosomes).

Nucleolus is always associated with the secondary constrictions of sat-chromosomes. Therefore, secondary constrictions are also called ***nucleolus organiser regions*** (*NOR*) and sat-chromosomes are often referred to as *nucleolus organiser chromosomes* (NOC). NOR of each sat-chromosome contains several hundred copies of the gene coding for ribosomal (rRNA).

Chromomere
- Beadlike structures in meiotic prophase (especially, pachytene) chromosomes.
- Distribution in the chromosomes highly characteristic.
- Form bands in giant chromosomes.
- Chromatin more condensed in chromomeres.

3.4.5. Chromomere

In some species, *e.g.,* maize, amphibia etc., chromosomes during the first prophase of meiosis (more particularly, during pachytene), show small bead-like structures called ***chromomeres.*** The distribution of chromomeres in a chromosome is highly characteristic and constant, the pattern of distribution being different for different chromosomes. *Homologous chromosomes always show an identical pattern of chromomere distribution. Chromomeres are most clearly visible in the dipteran giant salivary gland chromosomes as dark staining bands.*

Chromomeres present in a single chromosome show a considerable variation in size. Once it was believed that genes were located within chromomeres and that one chromomere may possibly represent a single gene. But available evidence does not

support this idea. Chromatin appears to be present in a more condensed state in chromomeres than in neighbouring regions of the chromosomes. This may account for both their larger diameter and greater stainability.

Karyotype
- Arrangement of somatic chromosomes in a descending order of size.
- Each chromosome is given a serial number.
- Characteristic for a species.

Idiotype
- Chromosome diagram arranged in a descending order of size.
- Usually haploid complement depicted.
- Yields the same information as karyotype.

3.5. KARYOTYPE

The general morphology (*i.e.,* the size of chromosomes, the position of centromeres, the presence of secondary constrictions and the size of satellite bodies) of the somatic chromosome complement of an individual constitutes its ***karyotype.*** Ordinarily, karyotypes are presented by arranging the chromosomes of somatic complement in the descending order of size keeping their centromeres in a straight line. Thus the longest chromosome is placed on the extreme left and the smallest one on the extreme right (Fig. 3.6). The sex chromosomes usually are placed in their appropriate positions according to their size, and are marked as X and Y.

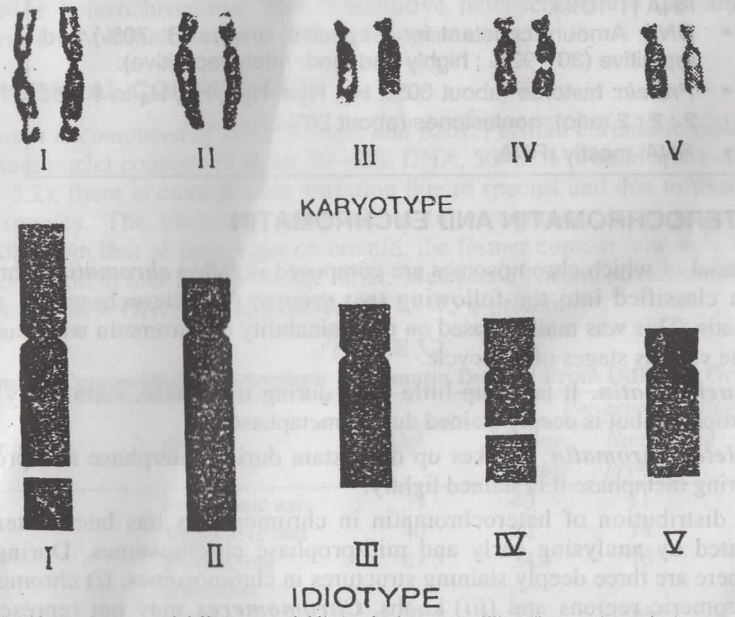


Fig. 3.6. Karyotype and idiotype of *Vicia hajastana.* The five pairs of chromosomes are arranged according to their length (in the descending order) both in the karyotype and the idiotype (only the haploid complement is depicted here). The chromosomes I and IV have secondary constrictions; the satellites are towards the bottom.

All the normal members of a species have an identical karyotype. Therefore, karyotype of a normal somatic cell of a normal individual represents the karyotype of the concerned species. Each chromosome of the karyotype is designated by a serial number according to its position in the karyotype.

3.6. IDIOTYPE

The karyotype of a species may be represented diagrammatically (in contrast to the actual photographs of chromosomes in karyotype) showing all the morphological features of the chromosomes; such a diagram is known as ***idiotype*** (Fig. 3.6). Ordinarily, idiotypes are prepared for the haploid chromosome complement of a species. An idiotype gives the same amount of information as does karyotype.

Chromatin

- Material constituting chromosomes.
- *Types*: euchromatin, heterochromatin.
- *Euchromatin*: faint stain during interphase, deep stain during metaphase; genetically active.
- *Hetrochromatin*: reverse of euchromatin; genetically much less active.
  - (*i*) Facultative: euchromatin changed to heterochromatic state.
  - (*ii*) *Constitutive*: always in heterochromatic state; distribution characteristic; mainly in centromeric and telomeric regions.
- *Chemical composition*: DNA (30 - 40%), protein (50 - 65%), RNA (1-10%).
- *DNA*: Amount constant for a species; unique (8 -70%) and repetitive (30 - 92% ; highly and moderately repetitive).
- *Protein*: histores (about 80%; H_1, H_2a, H_2b, H_3, H_4 in 1 : 2 : 2 : 2 : 2 ratio); nonhistones (about 20%).
- *RNA*: mostly rRNA.

3.7. HETEROCHROMATIN AND EUCHROMATIN

The material of which chromosomes are composed is called ***chromatin***. Chromatin has been classified into the following two groups: (1) heterochromatin and (2) euchromatin. This was mainly based on the stainability of chromatin with basic dyes during the various stages of cell cycle.

1. ***Euchromatin.*** It takes up little stain during interphase, stains only lightly during prophase, but is deeply stained during metaphase.

2. ***Heterochromatin***. It takes up deep stain during interphase and prophase, while during metaphase it is stained lightly.

The distribution of heterochromatin in chromosomes has been extensively investigated by analysing early and mid-prophase chromosomes. During these stages, there are three deeply staining structures in chromosomes: (*i*) chromomeres, (*ii*) centromeric regions and (*iii*) knobs. ***Chromomeres*** may not represent true heterochromatin. ***Centromeric regions*** invariably contain heterochromatin. ***Knobs*** are spherical structures, usually several times the diameter of the concerned chromosomes. They are present in certain chromosomes of some species, *e.g.*, maize. Where present, knobs serve as valuable chromosome markers.

Heterochromatin is classified into the following two groups: (*i*) constitutive and (*ii*) facultative. ***Constitutive heterochromatin*** remains permanently in the heterochromatic state, *i.e.*, it does not revert to the euchromatic state, *e.g.*, centromeric regions. In contrast, ***facultative heterochromatin*** is essentially euchromatin that has undergone heterochromatinization (conversion into heterochromatin). Heterochromatinization may involve a segment of a chromosome, a whole chromosome or one whole haploid set of chromosomes.

The distribution of constitutive heterochromatin within a chromosome is highly specific and characteristic. Special chromosome banding techniques have been developed to analyse heterochromatin distribution in chromosomes. *In general, heterochromatin is found in the centromeric and telomeric regions, but other regions may also contain heterochromatin. Heterochromatic chromosomes and chromosome regions ordinarily replicate later than do euchromatic regions.*

Genetic studies have revealed that ***constitutive heterochromatin is largely genetically inactive, while almost all of the genes are located in the euchromatic regions of chromosomes. However, heterochromatin may not be completely devoid of genes or genetic functions.***

It has been postulated that euchromatin and heterochromatin only represent different states of condensation of chromatin. *Heterochromatin is highly condensed during interphase and prophase, but is less condensed during metaphase, while euchromatin shows exactly the opposite pattern.* Electron micrographs of chromosomes tend to confirm this view. The genetic inactivity of heterochromatic regions is suggested to be due to their highly condensed state since highly condensed chromatin is generally not available for transcription. This is certainly the case for facultative heterochromatin. But constitutive heterochromatin contains highly repetitive DNA, which seems to contain few genes.

3.8. CHEMICAL COMPOSITION

Chromatin is composed of DNA, protein and RNA. Purified chromatin isolated from interphase nuclei consists of about 30–40% DNA, 50–65% protein and 1–10% RNA (Table 3.2); there is considerable variation due to species and due to tissues of the same species. The chemical composition of metaphase chromosomes differs markedly from that of interphase chromatin; the former contain relatively less DNA and more protein and RNA than the latter. Metaphase chromosomes contain, on an average, 15-20% DNA, 10–15% RNA and 65–75% protein.

TABLE 3.2

Chemical Composition of Interphase Chromatin Derived From Different Organisms

Organism	*Organ*	*Per cent constituent*			
		DNA	*Histone proteins*	*Nonhistone proteins*	*RNA*
Pea	Embryonic axis	39	40	11	10
	Vegetative bud	40	52	14	14
	Developing cotyledon	43.5	34.5	16	6
Rat	Liver	37	37	25	1
Calf	Thymus	40	46	13.5	0.5
Sea urchin	Larva	33	29	35	3
Human	He La cells	35.5	36	25	3.5

3.8.1. DNA

The amount of DNA present in normal (disomic) somatic cells of a species is constant for that species. Any variation in DNA from this value is strictly correlated with a corresponding variation (in number/size) at the chromosome level. Gametes of a species contain only half the amount of DNA present in its somatic cells. ***DNA is the only component of chromosomes the amount of which exhibits a constant relationship with the stage of chromosome replication and the increase or decrease in chromosome number.*** The DNA found in chromosomes is mainly of two types (*i*) unique DNA and (*ii*) repetitive DNA.

3.8.1.1. Unique DNA

Unique DNA consists of those DNA segments whose base sequences are present only in a single copy per genome, *i.e.*, their base sequences are not repeated in the genome. All the structural genes, except some, *e.g.*, genes for histones, ribosomal RNA, storage proteins etc., are made up of unique DNA. The proportion of unique DNA in cells of various species may vary from 70% (man) to only 8% (rye).

3.8.1.2. Repetitive DNA

Repetitive DNA base sequences are few to several hundred base pairs (bp) long and are present in several to a million copies per genome. The proportion of repetitive DNA in the genome varies from one species to the other; human genome has only 30% repetitive DNA, while in rye it constitutes about 92% of the genome. Repetitive DNA sequences may be either (*i*) highly repetitive or (*ii*) moderately repetitive.

Highly repetitive DNA sequences are present in more than 10^5 and upto several million copies per genome. *Highly repetitive sequences are located in heterochromatic regions of chromosomes (mainly the centromeric and telomeric regions).* ***Moderately repetitive DNA*** sequences are present in two to less than 10^5 copies per genome, and generally constitute 20–80% of the genome. *The moderately repetitive fraction includes genes for histones, ribosomal RNA and ribosomal proteins, and several types of transposable elements or nomadic sequences.* ***Transposable elements*** are capable of changing their location within the same chromosome or even migrate to a different chromosome. The function of these elements, either in evolution or in the organism in which they are found, is not quite clear. *Moderately repetitive DNA is distributed throughout the genome at variable intervals.*

3.8.2. Protein

Proteins associated with chromosomes may be classified into two broad groups: (1) basic proteins or histones and (2) nonhistone proteins. Nonhistone proteins are mainly acidic in nature, but a small fraction of these proteins may be neutral.

3.8.2.1. Histones

These constitute about 80% of the total chromosomal proteins. They are present in an almost 1: 1 ratio with DNA (weight/weight; Table 3.2). The molecular weight of histones ranges from 10,000 to 30,000 and *they are completely devoid of tryptophan. Histones are a highly heterogeneous class of proteins.* They are separable into 5 different fractions through electrophoresis. These fractions are

designated as H_1, H_2a, H_2b, H_3 and H_4 after Lewin (1975). *Fraction H_1 is lysine-rich, H_2a and H_2b are slightly lysine-rich, while H_3 and H_4 are arginine-rich.*

The approximate molar ratio of these fractions is $1H_1 : 2H_2a : 2H_2b : 2H_3 : 2H_4$. *Histones isolated from different organisms show a very high degree of similarity in four of the five histone fractions (H_2a, H_2b, H_3 and H_4), indicating that they have been highly conserved during evolution.* This suggests that these histones play a primary function in chromosome organisation. H_1 fraction is the most easily separated from chromatin. It has been suggested that fractions H_2a, H_2b, H_3 and H_4 are involved in the structural organisation of chromatin fibers, while fraction H_1 holds together the folded chromatin fibers of chromosomes.

3.8.2.2. Nonhistone Proteins

These make up about 20% of the total chromosome mass. But their amount is variable, and there is no definite ratio between the amounts of DNA and of nonhistone proteins present in chromosomes. There may be 12 to more than 20 different types of nonhistone proteins. *These proteins show variation from one species to the other and even in different tissues of the same organism. Clearly, they have not been as conserved during evolution as the histone proteins. This class of proteins includes many important enzymes, such as, DNA and RNA polymerases etc.*

3.9. CHROMOSOME STRUCTURE

One micron of a metaphase chromosome contains about 8,000 μ *of DNA double helix.* Thus any model for chromosome structure must be able to account for, among other things, the packaging of such a long DNA molecule into such a small segment of metaphase chromosome.

One of the earliest models of chromosome structure, based on light microscope studies, is presented in Fig. 3.7. According to this model, the basic component of chromosome structure is ***chromonema*** (pl., chromonemata), which is composed of chromatin and contains genes. The variation in chromosome length and thickness was proposed to be due to coiling and uncoiling of chromonemata. Each chromatid of a chromosome may contain two or more chromonemata, which run across through the centromere. In the case of metaphase chromosomes, chromonemata are surrounded by an amorphous ***matrix*** the outer side of which is enclosed in a membrane called ***pellicle.*** However, ***electron micrographs of metaphase chromosomes do not show any evidence for the existence of either matrix or pellicle. Clearly, this model of chromosome structure is inadequate and of historical interest only.***

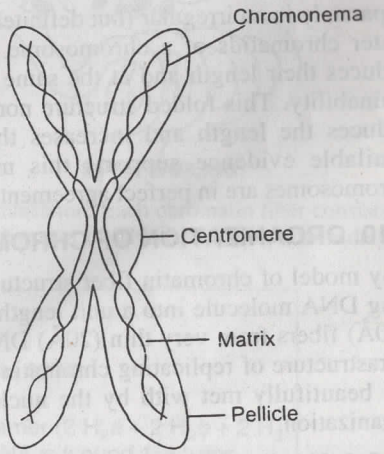


Fig. 3.7. A model of chromosome structure proposed on the basis of the studies with light microscope. (*This model is of historical interest only.*)

Several models of chromosome structure have been proposed from time to time based on various types of data on chromosomes. More recent models are mainly based on data from biochemical and electron microscope (EM) studies. *Studies of both intact and sectioned metaphase and interphase chromosomes have revealed that 300 Å diameter chromatin fibers are the basic units of chromosome structure*. The organisation of such chromatin fibers in chromosomes is best explained by the folded-fiber model of chromosome structure.

Folded-Fiber Model
- Chromosomes made up of 300Å chromatin fibers.
- One chromatid has a single chromatin fiber.
- Chromatin fiber replicates during *S* phase.
- Chromatin fibers fold in various ways and then supercoil to yield metaphase chromosomes.

3.9.1. Folded-Fiber Model

This model was proposed by DuPraw in 1965 and is widely accepted. According to this model, chromosomes are made up of chromatin fibers of about 230 Å diameter. This chromatin fiber is produced by coiling of a single DNA double helix; the coils of DNA molecule are stabilized by proteins. *Each chromatid contains a single enormously long chromatin fiber, which folds in various ways to yield the metaphase chromosome structure.*

The DNA of chromatin fiber replicates during interphase producing two sister chromatin fibers. But the fiber remains unreplicated in the centromeric region so that the two sister fibers remain joined in this region (Fig. 3.8). Subsequently, the chromatin fiber undergoes replication in the centromeric region as well so that the sister chromatin fibers become separated in this region also.

During cell division, the two sister chromatin fibers undergo extensive folding separately in an irregular (but definitely reproducible) manner to give rise to the two sister chromatids of a chromosome. Folding of the chromatin fibers drastically reduces their length and at the same time markedly increases their thickness and stainability. This folded structure normally undergoes supercoiling, which further reduces the length and increases the thickness of chromosomes. Most of the available evidence supports this model; electron micrographs of metaphase chromosomes are in perfect agreement with the model (Fig. 3.4).

3.10. ORGANIZATION OF CHROMATIN FIBERS

Any model of chromatin fiber structure has to account for (*i*) packaging of a very long DNA molecule into a unit length of fiber; (*ii*) production of very thick (230–300Å) fibers from very thin (20Å) DNA molecules, and (*iii*) the beads-on-a-string ultrastructure of replicating chromatin fibers (Fig. 3.9). These and other conditions are beautifully met with by the nucleosome–solenoid model of chromatin fiber organization.

3.10.1. Nucleosome-Solenoid Model

This model was proposed by Kornberg and Thomas in 1974 and is the most widely accepted. According to this model, the chromatin is composed of a repeating unit called nucleosome. One ***complete nucleosome*** consists of (1) a nucleosome core, (2) linker DNA, (3) an average of one molecule of histone H_1 and (4) the other associated chromosomal proteins.

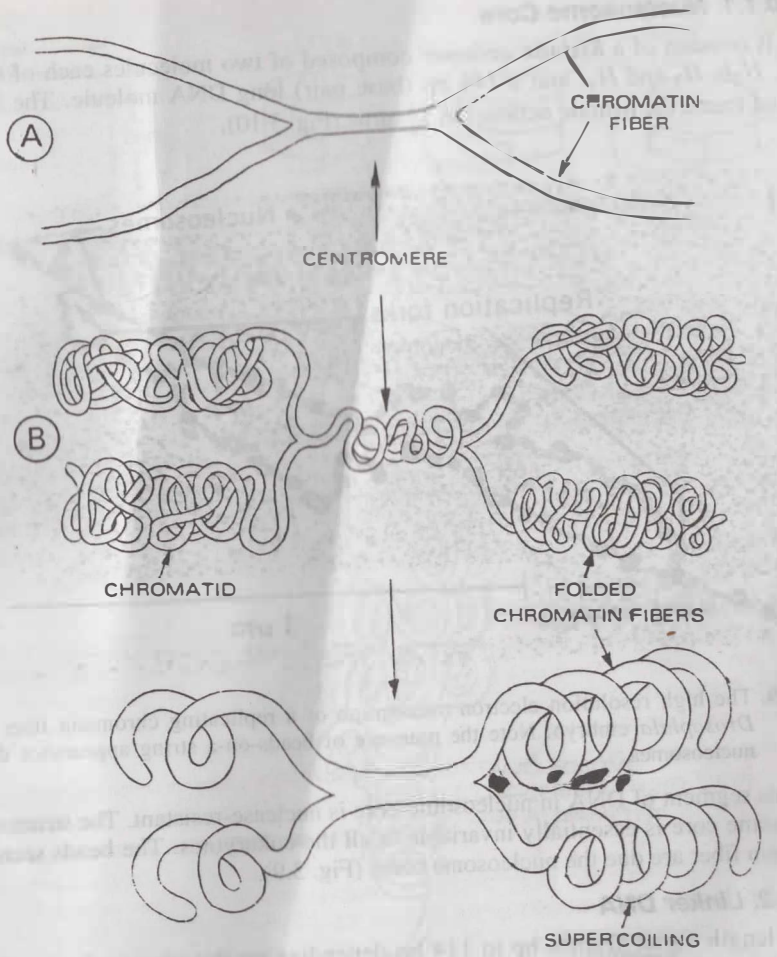


Fig. 3.8. The folded-fiber model of chromosome organisation. Each chromatin fiber consists of one DNA molecule and has the average diameter of 230 Å. This model is widely accepted.

Nucleosome-Solenoid Model
- Model of chromatin fiber structure.
- Proposed by Kornberg and Thomas in 1974.
- Consists of:
  - (*i*) *Nucleosome core* : histone octamer (2 H_2a + 2 H_2b + 2 H_3 + 2 H_4) around which 146 bp DNA is wound 1¾ turns.
  - (*ii*) *Linker DNA*: 8 -114 bp depending on species; links two neighbouring nucleosome cores; DNA nuclease-sensitive-yields the beads-on-a-string chromatin fiber
- *Nuclosome fiber* : nucleosome cores pack to yield 100 Å thick fiber.
- *Solenoid* : nucleosome fiber supercoils to yields 300 Å fiber.
- Universally accepted.

3.10.1.1. Nucleosome Core

It consists of a ***histone octamer*** composed of two molecules each of histones H_2a, H_2b, H_3 and H_4, and a 146 bp (base pair) long DNA moleule. The DNA is wound round the histone octamer in $1\frac{3}{4}$ turns (Fig. 3·10).

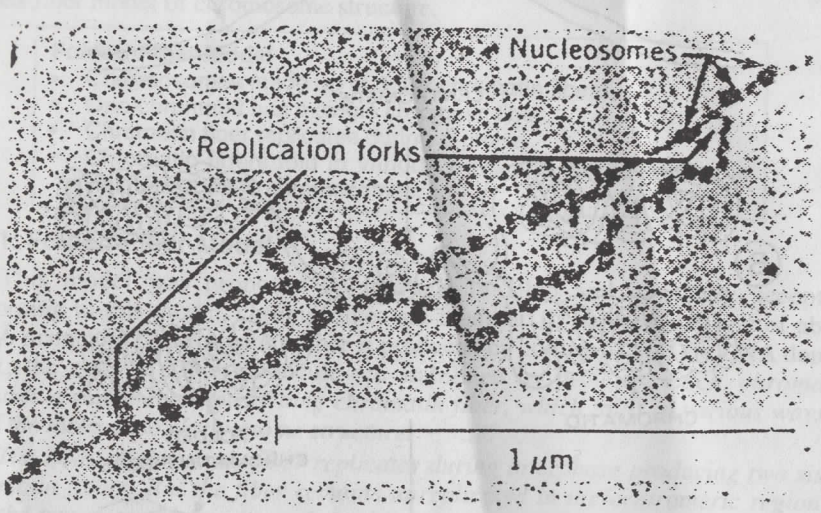


Fig. 3.9. The high resolution electron micrograph of a replicating chromatin fiber from *Drosophila* embryo. Note the presence of beads-on-a string appearance due to nucleosomes.

This segment of DNA in nucleosome-core is nuclease-resistant. The structure-of nucleosome core is essentially invariable in all the eukaryotes. The beads seen in a chromatin fiber are due the nucleosome cores (Fig. 3.9).

3.10.1.2. Linker DNA

Its length varies from 8 bp to 114 bp depending on the species. Linker DNA forms the string part of the beads-on-a string chromatin fiber, and is nuclease-susceptible. Both linker DNA and nucleosome core are associated with other chromosomal proteins. Each nucleosome contains, on an average, one molecule of H_1 histone. Some studies suggest that H_1 is associated on the outside of each nucleosome core, and that one H_1 molecule stabilizes about 166 bp long DNA molecule. Other studies, however, suggest that the molecules of H_1 histone are involved in stabilizing the supercoils of nucleosome chromatin fibers.

In native chromatin the beads are about 110 Å in diameter, Ca. 60 Å high, and ellipsoidal in shape. Under some conditions, nucleosomes pack together without any detectable linker DNA between them; this produces a 100 Å thick chromatin fiber called ***nucleosome fiber*** (Fig. 310).

The nucleosome fiber may supercoil to give rise to the 300 Å chromatin fiber seen in electron micrographs of metaphase chromosomes. The supercoiled nucleosome fiber is called ***solenoid.*** The nucleosome model of chromatin fiber structure is consistent with almost all of the evidence accumulated so far.

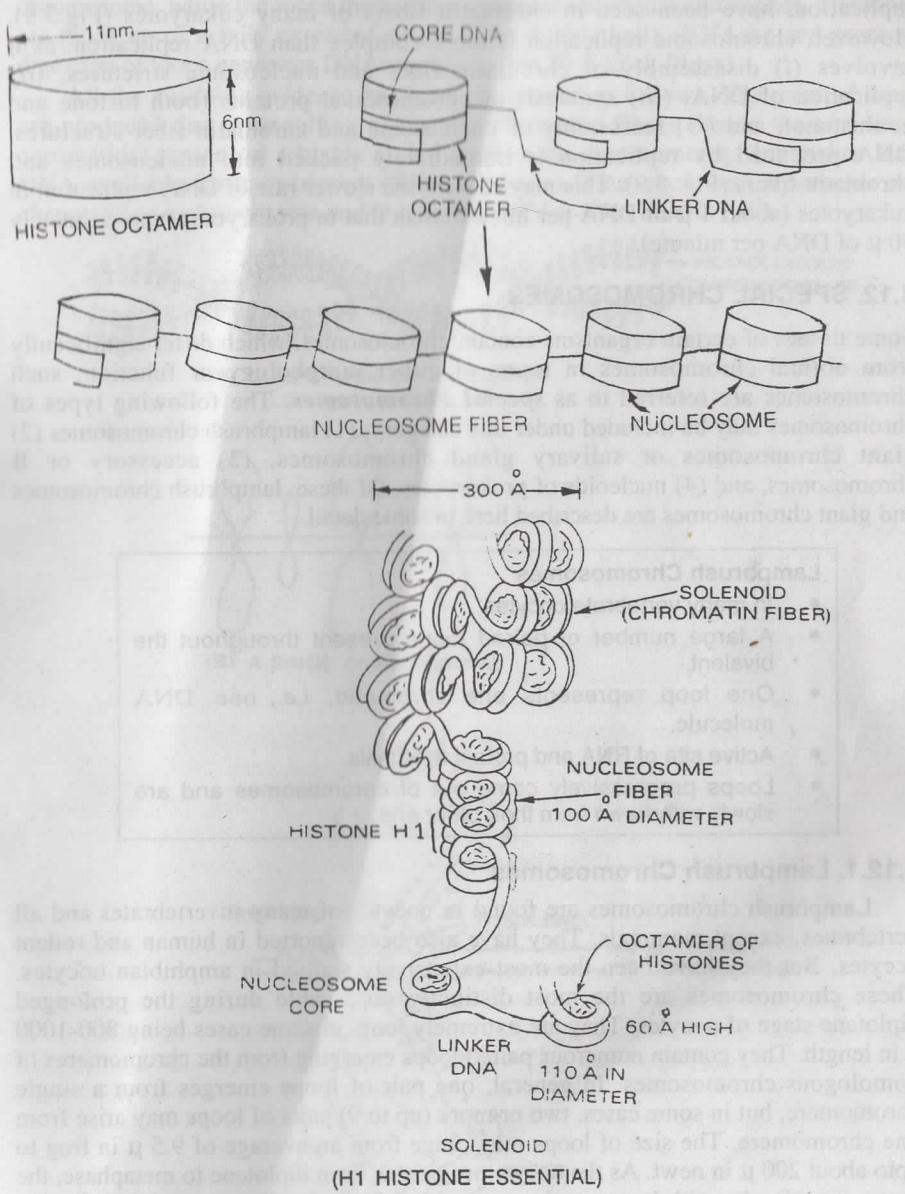


Fig. 3.10. The nucleosome-solenoid model of chromatin fiber organisation. (***This model is widely accepted***).

3.11. REPLICATION OF CHROMOSOMES

The mode of replication of eukaryotic chromosomes, like that of prokaryotic ones, is semiconservative, which is the same as that for DNA replication. This was first demonstrated in 1956 by J.H. Taylor and coworkers by radiolabelling the chromosomes of *Vicia faba.* Replication forks, expected in a semiconservative

replication, have been seen in chromatin fibers of many eukaryotes (Fig.3.9). However, chromosome replication is more complex than DNA replication, as it involves (*i*) disassembly of chromatin fiber and nucleosome strictures, (*ii*) replication of DNA, (*iii*) synthesis of chromosomal proteins (both histone and nonhistone), and (*iv*) reassembly of nucleosome and chromatin fiber structures. DNA produced by replication is immediately packed into nucleosomes and chromatin fibers (Fig. 3.9). This may explain the slower rate of DNA replication in eukaryotes (about 1 μ of DNA per minute) than that in prokaryotes (approximately 30 μ of DNA per minute).

3.12. SPECIAL CHROMOSOMES

Some tissues of certain organisms contain chromosomes, which differ significantly from normal chromosomes in terms of either morphology or function; such chromosomes are referred to as ***special chromosomes.*** The following types of chromosomes may be included under this category: (1) lampbrush chromosomes (2) giant chromosomes or salivary gland chromosomes, (3) accessory or B chromosomes, and (4) nucleoids of prokaryotes. Of these, lampbrush chromosomes and giant chromosomes are described here in some detail.

Lampbrush Chromosomes

- In many vertebrate oocytes.
- A large number of paired loops present throughout the bivalent.
- One loop represents one chromatid, *i.e.*, one DNA molecule.
- Active site of RNA and protein synthesis.
- Loops progressively come out of chromosomes and are slowly withdrawn from their other end.

3.12.1. Lampbrush Chromosomes

Lampbrush chromosomes are found in oocytes of many invertebrates and all vertebrates, except mammals. They have also been reported in human and rodent occytes. But they have been the most extensively studied in amphibian oocytes. These chromosomes are the most distinctly observable during the prolonged diplotene stage of oocytes. They are extremely long, in some cases being 800-1000 μ in length. They contain numerous paired loops emerging from the chromomeres of homologous chromosomes. In general, one pair of loops emerges from a single chromomere, but in some cases, two or more (up to 9) pairs of loops may arise from one chromomere. The size of loops may range from an average of 9.5 μ in frog to upto about 200 μ in newt. As the oocyte progresses from diplotene to metaphase, the loops are slowly withdrawn and reassembled into chromomeres. Metaphase I chromosomes of these organisms have perfectly normal morphology.

During diplotene, homologous chromosomes begin to separate from each other. Therefore, a lampbrush chromosome consists of two homologous chromosomes that are in contact with each other at several points (chiasmata). Each chromosome has several chromomeres distributed over its length. From each chromomere, a pair of loops emerges in the opposite directions vertical to the main chromosomal axis (Fig. 3.11A). The main chromosomal axis consists of the two sister chromatids of a

chromosome, while the loops themselves represent individual chromatids. The loops are made up of a fully extended single DNA double helix (RNAase and protease digestion of loops generates DNAase susceptible 20 Å thick fibers).

All the available evidence has been interpreted as follows. The pairs of loops are produced due to uncoiling of the two chromatin fibers (hence the two sister chromatids) present in a highly coiled state in the chromomeres; this makes their DNA available for transcription (RNA synthesis). Thus *each loop represents one chromatid of a chromosome and is composed of one DNA double helix.*

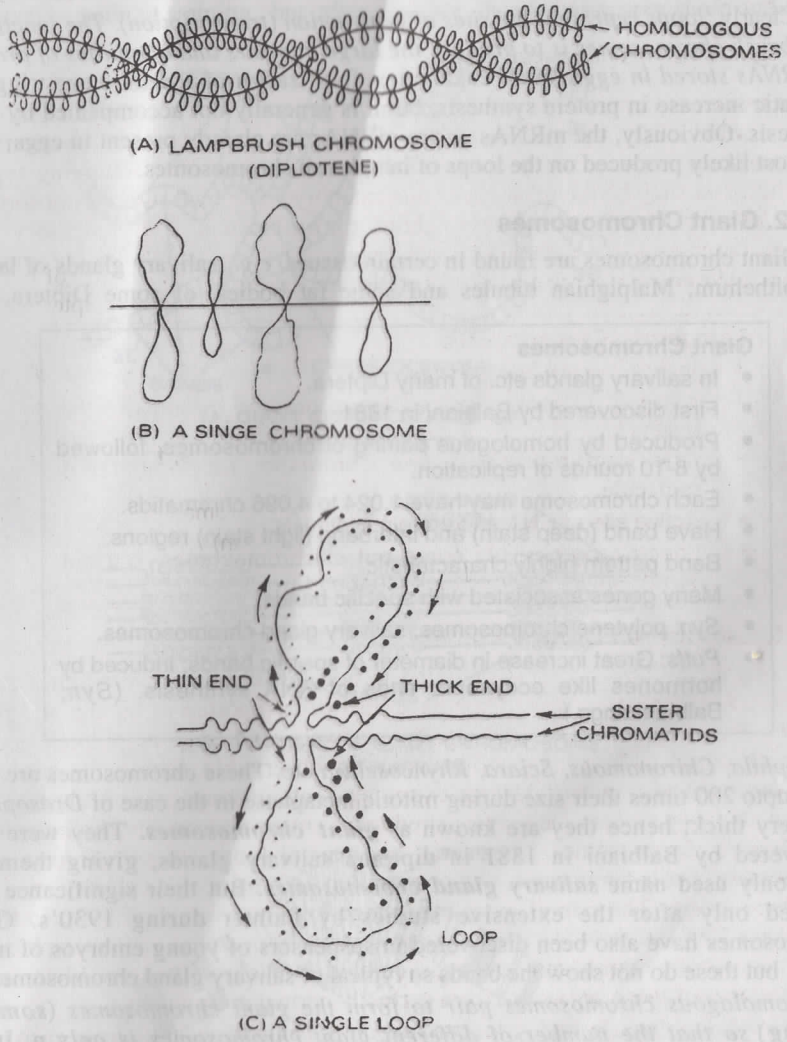


Fig. 3.11. Lampbrush chromosomes found in oocytes of most vertebrates. (A) A segment of a bivalent (the paired chromosome structure) during diplotene. (B) A magnified diagram of one homologue of the pair. (C) A highly magnified diagram of a single loop (the arrows show the direction of movement of the DNA molecule within the loop).

One end of each loop is markedly thinner (***thin end***) than the other end (***thick end***). *There is extensive RNA synthesis at the thin end of loops, while there is little or no RNA synthesis at the thick end. The chromatin fiber of the chromomere is progressively uncoiled toward the thin end of a loop. The DNA at the thick end of a loop is progressively withdrawn and reassembled into the chromomere.* Thus the DNA axis of loop moves progressively from the thin end toward the thick end of loops. *It takes about* 10 *days for a point to move from the thin end to the thick end of the loop.*

Clearly, *loops represent the sites of gene action (transcription). The function of lampbrush chromosomes is to produce the large numbers and quantities of proteins and RNAs stored in eggs.* In all eukaryotes, fertilization of a mature egg triggers a dramatic increase in protein synthesis, but it is generally not accompanied by RNA synthesis. Obviously, the mRNAs or pre-mRNAs are already present in eggs; these are most likely produced on the loops of lampbrush chromosomes.

3.12.2. Giant Chromosomes

Giant chromosomes are found in certain tissues, *e.g.*, salivary glands of larvae, gut epithelium, Malpighian tubules and some fat bodies, of some Diptera, *e.g.*,

Giant Chromosomes

- In salivary glands etc. of many Diptera.
- First discovered by Balbiani in 1881.
- Produced by homologous pairing of chromosomes, followed by 8-10 rounds of replication.
- Each chromosome may have 1,024 to 4,096 chromatids.
- Have band (deep stain) and interband (light stain) regions.
- Band pattern highly characteristic.
- Many genes associated with specific bands.
- *Syn*: polytene chromosomes, salivary gland chromosomes.
- *Puffs*: Great increase in diameter of specific bands; induced by hormones like ecdysone; sites of RNA synthesis. (*Syn*; Balbiani rings.)

Drosophila, Chironomous, Sciara, Rhyncosciara etc. These chromosomes are very long (upto 200 times their size during mitotic metaphase in the case of *Drosophila*) and very thick; hence they are known as ***giant chromosomes.*** They were first discovered by Balbiani in 1881 in dipteran salivary glands, giving them the commonly used name ***salivary gland chromosomes***. But their significance was realized only after the extensive studies by Painter during 1930's. Giant chromosomes have also been discovered in suspensors of young embryos of many plants, but these do not show the bands so typical of salivary gland chromosomes.

Homologous chromosomes pair to form the giant chromosomes (***somatic pairing***) *so that the number of different giant chromosomes is only* n. In *D, melanogaster* salivary glands, giant chromosomes radiate as five long and one short arms from a deeply staining and more or less amorphous structure, called chromocentre. The ***chromocentre*** is formed by the fusion of centromeric regions of all the chromosomes and, in males, of the entire Y chromosome. The short arm radiating from the chromocentre represents chromosome IV, one of the long arms is

due to the X chromosome, while the remaining four long arms represent the arms of chromosomes II and III (Fig. 3.12A).

Slides of *Drosophila* giant chromosomes are prepared by squashing in acetocarmine the salivary glands dissected out from the 3rd instar larvae. The total length of *D. melanogaster* giant chromosomes is about 2,000 μ, while its metaphase (mitosis) chromosomes measure only 7.5 μ. Giant chromosomes are made up of several dark staining regions called ***bands*** separated by relatively light or nonstaining ***interband regions.*** The bands in *Drosophila* giant chromosomes are visible even without staining, but after staining they become very sharp and clear.

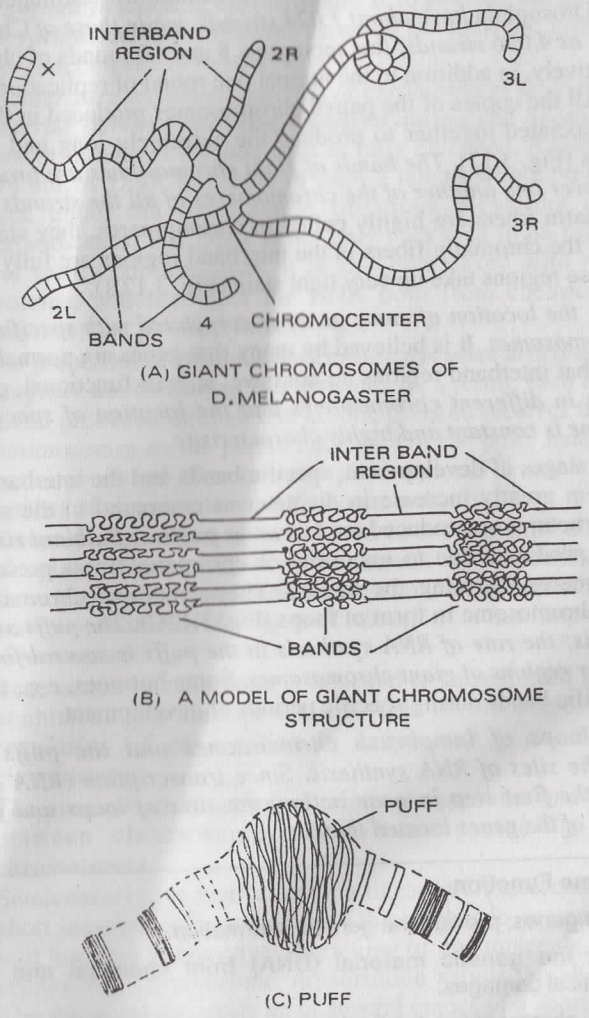


Fig. 3.12. Giant chromosomes of *D. melanogaster* salivary glands. (A) The entire giant chromosome. (B) A model of the structure of giant chromosomes. (C) A puff in a band of a giant chromosome.

The total number of bands in the four chromosomes of *D. melanogaster* is estimated as more than 5,000. Some of these bands are as thick as 0.5 μ, while some may be only 0.05 μ thick (visible only in electron micrographs).

All the available evidence clearly shows that each giant chromosome is composed of numerous strands, each strand representing one chromatid. Therefore, these chromosomes are also known as ***polytene chromosomes,*** and the condition is referred to as ***polyteny***. The numerous strands of these chromosomes are produced due to repeated replication of the paired chromosomes without any nuclear or cell division (***endoreduplication***) so that the number of strands (chromatids) in a chromosome doubles after every round of DNA replication. It is estimated that *giant chromosomess of* Drosophila *have about 1,024 strands, while those of* Chironomous *may have as many as 4,096 strands;* this represents 8 and 10 rounds of chromosome replication, respectively, in addition to the normal one round of replication following the cell division. All the copies of the paired chromosomes produced in this manner remain closely associated together to produce the extremely long and very thick giant chromosomes (Fig. 3.12). *The bands of giant chromosomes are produced as a result of stacking over one another of the chromomeres of all the strands present in them.* Since chromatin fibers are highly coiled in chromomeres, they stain deeply. On the other hand, the chromatin fibers in the interband regions are fully extended, and, as a result, these regions take up very light stain (Fig. 3.12B).

In Drosophila, *the location of many genes is correlated with specific bands in the concerned chromosomes.* It is believed by many that genes are normally located within bands, and that interband regions do not have, at least functional, genes. *The pattern of banding in different chromosomes and the location of specific bands within a chromosome is constant and highly characteristic.*

During certain stages of development, specific bands and the interband regions associated with them greatly increase in diameter as compared to the rest of the chromosome; the structures so produced are known as ***puffs*** or ***Balbiani rings.*** Puffs are believed to be produced due to uncoiling of chromatin fibers present in the concerned chromomeres (forming the bands). These uncoiled chromatin fibers extend outside the chromosome in form of loops (Fig. 3.12C). *The puffs are sites of active RNA synthesis; the rate of RNA synthesis in the puffs is several-fold higher than that in the other regions of giant chromosomes.* Some hormoes, *e.g.,* ecdysone, induce puffs in specific bands during specific periods of development.

Clearly, the loops of lampbrush chromosomes and the puffs of giant chromosomes are the sites of RNA synthesis. Since transcription (RNA synthesis based on DNA) is the first step in gene action, the sites of loops and puffs are indicative of activity of the genes located therein.

Chromosome Function

- Contain genes; provide the genetic information.
- Protect the genetic material (DNA) from chemical and mechanical damages.
- Precise chromosome distribution due to centromeric function.
- Chromosome stability due to telomeres.
- Regulation of gene function.

3.13. FUNCTIONS OF CHROMOSOMES

Although chromosomes were first described in 1875, their role in heredity was suggested independently by Sutton and Boveri in 1902. The chromosome theory of inheritance was slow in gaining acceptance. But *it is now universally accepted that genes are located in chromosomes.* This and various other functions of chromosomes may be summarised as under.

1. *It is universally accepted that DNA is the genetic material, and that in eukaryotes almost all the DNA is present in chromosomes. Thus the most important function of chromosomes is to* ***provide the genetic information*** *for various cellular functions essential for survival, growth, development, reproduction etc. of organisms.*
2. *Another very important function of chromosomes is to* ***protect the genetic material*** *(DNA) from being damaged during cell division.* If all the DNA present in human somatic cells is fully extended, it would measure about 174 cm. A DNA molecule of such a length would be prone to breakage even under the most gentle situations prevailing in a cell. But in chromosomes, this DNA is not only packaged in 46 relatively very small and sturdy chromosomes, it is also complexed with histones and other proteins, which protect the DNA both from chemical (*e.g.,* nuclease enzymes) and physical forces.
3. The properties of chromosomes ensure a ***precise distribution of DNA to the daughter nuclei during cell division.*** As a result, each of the two daughter nuclei produced through one mitotic division receives the same number of chromosomes as the parent nucleus. On the other hand, each of the four nuclei produced after one complete meiotic division gets only half the number of chromosomes as the parent nucleus.
4. Chromosome movement is believed to occur mainly due to the shortening of spindle fibers. These fibers attach to the centromeric regions of chromosomes. Thus ensuring ***a regular and directional movement of the genetic material*** *during cell division is an important function of chromosomes.*
5. Gene action in eukaryotes is believed to be regulated through histone and nonhistone proteins associated with chromosomes. *Participation in* ***regulation of gene action in eukaryotes*** *is another function of the chromosomes.*
6. The specialized structure of telomeres prevents end to end interactions between chromosomes. This ensures ***stability of the different chromosomes.***
7. Semiconservative replication of a linear DNA molecule inevitably leaves a short sequence at the two ends of this molecule in unreplicated state. This will lead to a progressive shortening of the molecule after every round of replication. Chromosome organization avoids this shortening as follows. The telomeres are made up of several copies of a short repeating sequence. The enzyme telomerase adds copies of this sequence at the chromosome ends. Therefore, the shortening that occurs due to replication is adequately compensated for by telomerase action, and the chromosome length is maintained.

QUESTIONS

1. What are chromosomes ? Describe the chromosome morphology as seen under light microscope with the help of suitable diagrams.
2. Describe the morphology, ultrastructure and function of centromere.
3. List the chemical constituents of chromosomes and give a detailed account of these constituents.
4. Describe the folded-fiber model of chromosomes with the help of suitable diagrams.
5. Describe in some detail the nucleosome-solenoid model of chromatin fiber organisation.
6. Describe chromosomes under the following headings:
   (*i*) chromosome number, (*ii*) chromosome size
   (*iii*) secondary constriction and satellite, (*iv*) chromomere, and
   (*v*) karyotype and idiotype.
7. List the various types of special chromosomes found in plants and animals. Describe in detail the structural organisation and function of any one of these chromosomes.
8. Write short notes on the following:
   (*i*) chromatid, (*ii*) centromere,
   (*iii*) karyotype, (*iv*) chromomere,
   (*v*) histone, (*vi*) repetitive DNA,
   (*vii*) folded-fiber model of chromosomes, (*viii*) nucleosome,
   (*ix*) chromosome replication, (*x*) functions of chromosomes,
   (*xi*) puffs, (*xii*) salivary gland chromosomes,
   (*xiii*) lampbrush chromosomes.

4

Cell Division I. Mitosis

4.1. CELL DIVISION

Cell is the structural and functional unit of (almost) all living organisms, and all cells originate through division of pre-existing cells. Bodies of all multicellular organisms are derived from unicellular zygotes through repeated divisions of the zygote and the cells derived through its division. *The division of chromosomes and cytoplasm of a cell into two daughter cells is known as* ***cell division.*** The cell that undergoes division is termed as ***parent cell,*** while the cells generated through the division of a cell are known as ***daughter cells.***

Cell Division

- Division of chromosomes (*karyokinesis*) and cytoplasm (*cytokinesis*) of one cell into two.
- *Parent cell* : the cell that divides.
- *Daughter cells* : cells obtained from division.
- *Functions* :
  (*i*) Production of numerous cells from a single cell, which:
  (*ii*) Allows growth, repair and differentiation of tissues/organs.
  (*iii*) Regulates volume / surface ratio of cells.
  (*iv*) Allows sexual and asexual reproduction, and evolution.
  (*v*) Smaller cells are less prone to damage.
- *Types* : mitosis and meiosis.

4.2. FUNCTIONS OF CELL DIVISION

1. *The main function of cell division* is to produce daughter cells from a single parent cell. As a result, repeated cell divisions generate a large number of cells. These cells function in:
   (*i*) growth and development of somatic tissues of organisms,
   (*ii*) regeneration of damaged tissues, and production of new organs and tissues,
   (*iii*) replacement of old organs and tissues,
   (*iv*) asexual reproduction
   (*v*) sexual reproduction.
2. Cell Division keeps the size of cells within a limited range. If the growth of an organism occurred only due to an enlargement of the zygote, it will be composed of a single cell. In such organisms, differentiation of specialised tissues/organs, like eyes, brain, gut etc., will be impossible.
3. An enlargement of a cell will lead to a relatively much more rapid increase in cell volume than in cell surface. A cell obtains its food and excretes its

waste products through its surface. Therefore, as the cell size increases, the quantum of its food requirement and that of excretion of waste products will also increase proportionately. As a consequence, a cell larger than a certain size is sure to die due to starvation and accumulation of cellular wastes.

4. A large cell is more likely to be damaged than a smaller cell. In view of these, cell division evolved and developed very early in organic evolution.
5. Finally, new cells cannot arise from pre-existing cells without cell division. Therefore, *cell division is a prerequisite for the continuity of life and for the evolution of various life forms.*

4.3. TYPES OF CELL DIVISION

There are mainly two types of cell division: (1) mitosis and (2) meiosis. The various events occurring during cell division may be grouped into two categories: (1) karyokinesis and, (2) cytokinesis.

1. ***Karyokinesis.*** It is the division of chromosomes of a cell into two groups; these give rise to the two daughter nuclei.

2. ***Cytokinesis.*** It is the division of cytoplasm of a cell into two halves to produce two daughter cells. Each cell ordinarily contains a single daughter nucleus. *As a rule, cytokinesis follows karyokinesis.*

4.4. MITOSIS

The term mitosis was first used by Fleming in 1882. *In mitosis, the two chromatids of each chromosome separate and move to the opposite poles of a cell. As a result, the two daughter nuclei produced by mitotic division of a nucleus are identical to the parent nucleus (both in terms of chromosome number and the kinds of chromosomes present).*

Mitosis

- Term coined by Fleming (1882).
- *Features* :
  - (*i*) The two chromatids of each chromosome move to opposite poles.
  - (*ii*) Produces two daughter cells.
  - (*iii*) Daughter cells are identical to the parent cell.
- *Occurrence* : Somatic cells, spores, reproductive tissues (until meiosis begins).
- *Functions :*
  - (*i*) growth and development,
  - (*ii*) regeneration of damaged tissues,
  - (*iii*) replacement of old tissues,
  - (*iv*) production of new organs,
  - (*v*) asexual reproduction,
  - (*vi*) gametogenesis in spores (in case of plants).
- *Stages* : (*i*) interphase (*ii*) prophase, (*iii*) metaphase, (*iv*) anaphase, and (*v*) telophase.

4.4.1. Occurrence

Somatic cells of all the organisms divide through mitosis. In addition, *cells in reproductive tissues undergo mitosis, upto just prior to the meiotic division. Nuclei of the spores, produced by meiosis, also divide by mitotic division.* Development of embryos from zygotes involves repeated mitotic division. In the initial stages, all the cells of a young embryo undergo cell division. But as the development of embryos progresses, cell division becomes confined to some specific organs and tissues.

In plants, mitosis is confined to the meristematic tissues of root and shoot tips, young leaves, flower buds and cambium. In adult animals, mitosis is ordinarily confined to the organs concerned with the production of blood cells, outer skin cells and gut epithelium. Healing of wounds and replacement of damaged organs in both plants and animals is based on mitotic cell division.

4.4.2. Importance of Mitosis

The chief functions of mitosis are growth of organisms and regeneration of damaged tissues. The various functions of mitosis may be summarised as follows.

1. *Production of large number of cells identical to the zygote in chromosome and gene contents. This enables growth and development of organisms.*
2. *Regeneration of damaged tissues and organs.*
3. Regular and continuous replacement of some old tissues, *e.g.*, gut epithelium, blood cells etc.
4. Production of new organs, *e.g.*, root and shoot branches in plants.
5. Production of progenies identical to the parent plant through *asexual reproduction.*
6. *In plants, production of the male and female gametes* through mitosis in micro- and mega-spores, respectively.

4.5. STAGES OF MITOSIS

The events in a mitotic cell division are grouped into five stages: (1) interphase, (2) prophase, (3) metaphase, (4) anaphase and (5) telophase (Figs. 4.1, 4.2). This classification is primarily for convenience in description and discussion. It does not indicate that mitotic cell division occurs in five discrete and distinct steps. In fact, *cell division is a continuous process in which a cell progresses gradually from one*

Interphase

- Chromosomes fully extended as chromatin fibers: not visible under light microscope.
- DNA, RNA, protein syntheses.
- *G1* (*first gap*): before *S* phase, RNA and protein syntheses.
- *S* (*DNA Synthesis*): after *G1*, chromosome replication; histone synthesis.
- *G2* (*second gap*): after *S*; considerable RNA and protein syntheses.
- DNA replication and cell growth essential for mitosis to start.
- *Replicon* : A *DNA* segment capable of replication; each bacterial chromosome has one, each enkaryotic chromosome has several replicons.

stage into another. The boundaries between any two stages are not at all clearly defined.

4.5.1. Interphase

Interphase stage occurs after telophase stage of the previous mitotic division and before the onset of prophase stage of the next one. During interphase, chromosomes are fully extended and uncoiled. They do not take up sufficient stain to be visible under light microscope. Nuclei are surrounded by a well defined nuclear envelope, and nucleoli are also present (Fig. 4.1A).

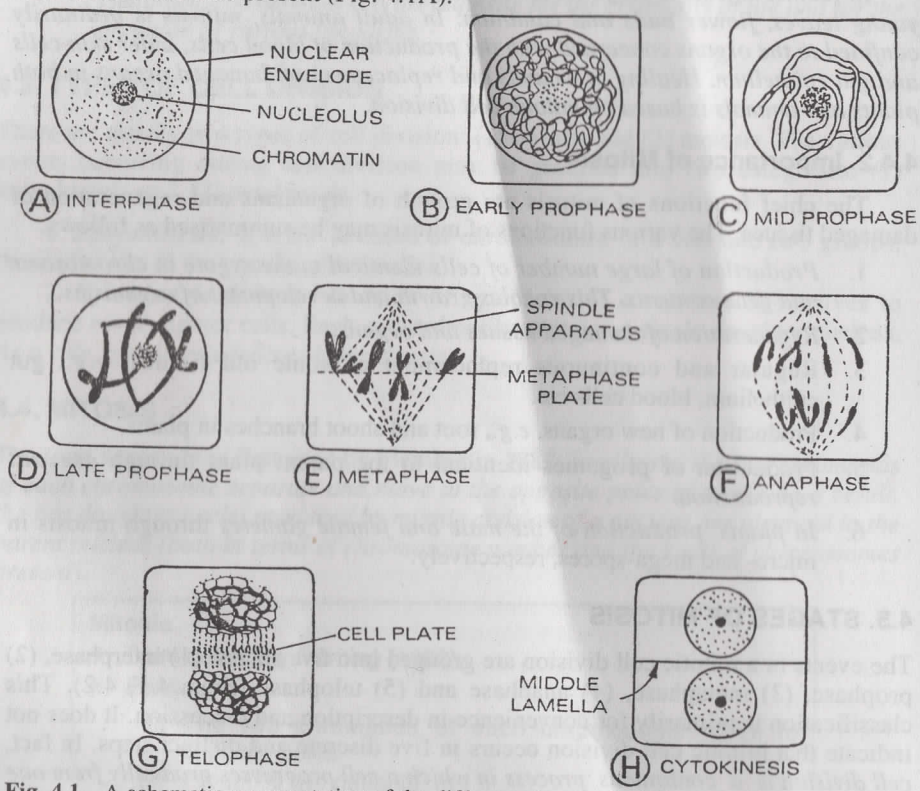


Fig. 4.1. A schematic representation of the different phases of mitotic cell division in a plant cell with $2n = 4$. The prophase stage has been subdivided into three substages on the basis of the degree of chromosome condensation.

Interphase is often not regarded as a stage of cell division. Therefore, early cytologists referred to it as ***resting stage.*** But *interphase is not a resting stage. Chromosome replication, and protein and RNA syntheses take place during this stage. These events are essential for the cell to enter mitotic division. Interphase is the longest stage of cell cycle* (section 4.6). All the nondividing and differentiated cells are present in this stage; such cells are said to be ***quiescent*** and are placed in G_0 stage to indicate their inability to divide. Interphase is divided into the following three substages: (1) G_1 (first gap), (2) S (synthesis, DNA) and (2) G_2 (second gap).

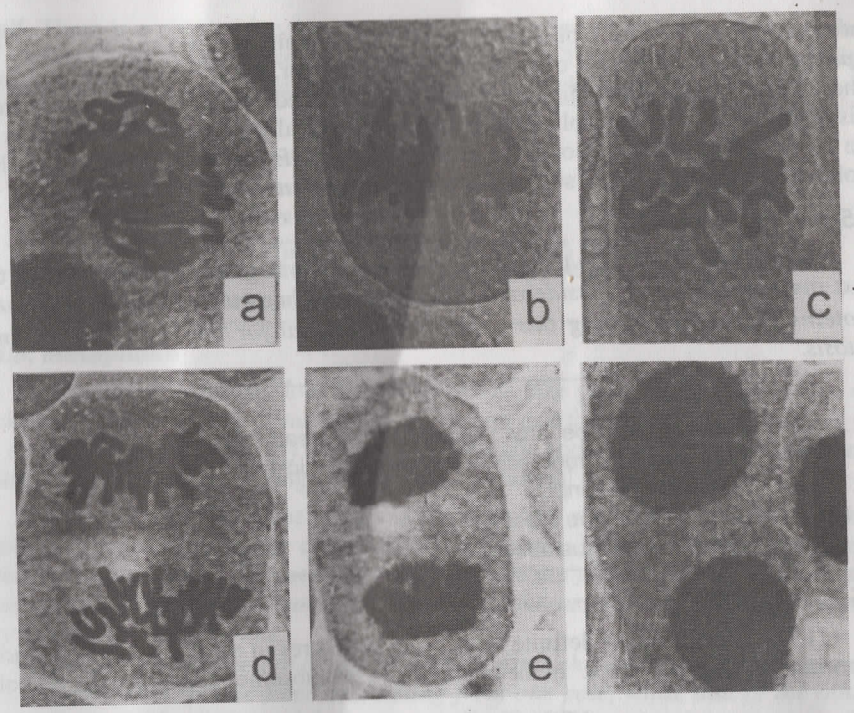


Fig. 4.2. The different stages of mitosis. (*a*) prophase, (*b*) metaphase (side view), (*c*) metahase polar view, (*d*) anaphase (late), (*e*) telophase (early), and (*f*) telophase (late). (Photographs courtesy of Prof. P.K. Sinha, Ranchi.)

4.5.1.1. G_1 Substage

It is the period between the beginning of interphase and the beginning of DNA synthesis (S phase). In some cells, G_1 may be very short or absent, while nondividing cells may be permanently in G_1. In cells preparing for call division, there is marked protein and RNA synthesis during G_1. The enzymes and substrates necessary for DNA synthesis during S phase are also synthesised during this phase.

4.5.1.2. S Phase

It is the period of interphase during which DNA synthesis takes place. As a result of this, chromosomes replicate during S phase. *Each chromosome consists of a single chromatid during telophase and G_1. DNA synthesis during S phase leads to the production of two sister chromatids from the single chromatid of each chromosome.* Therefore, at the and of S phase each chromosome is composed of two morphologically and genetically identical sister chromatids. RNA and protein syntheses are very low during S phase. But *almost all the histones of the nucleus are synthesised during this period. DNA synthesis is a prerequisite for the entry of a cell into mitosis, but it does not always lead to mitosis. Many cells may become arrested in the G_2 phase and may not divide at all.*

4.5.1.3. Replicon

A DNA segment capable of undergoing replication is known as a ***replicon***. The entire bacterial chromosome is a single replicon. But in eukaryotes, each

chromosome has several replicons (Fig. 4.3). Each replicon has a specific point, the ***replication initiation point*** or ***origin of replication*** (or, more simply, ***origin***), from where DNA replication is initiated. Replication proceeds in both the directions from this point. In eukaryotes, replication initiation is generally almost simultaneous in all the replicons of a chromosome. In prokaryotes like *E. coli*, the replicon also has replication termination site; such a site is called ***terminus***.

4.5.1.4. G_2 Phase

It is the period between the end of *S* phase and the beginning of prophase of the next division. *There is considerable RNA and protein synthesis during G_2. Some proteins synthesized during this period are essential for the entry of cells into mitosis.*

Prophase
- Chromosomes become visible; initially thread - like.
- They become progressively shorter and thicker.
- Nucleolus present.
- Nuclear envelope present.
- Spindle apparatus absent.

4.5.2. Prophase

(1) The appearance of definite thread-like structures in nuclei is the single most important event of prophase. In the beginning, chromosomes appear as a loose ball

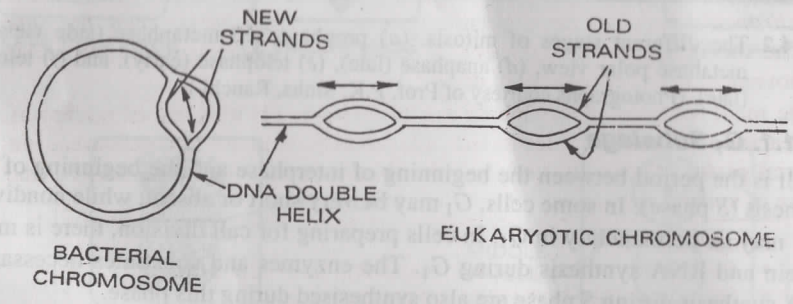


Fig. 4.3. Replication of the chromosomes of bacteria and eukaryotes. (A) replication of bacterial chromosome begins at one point and proceeds in both the directions. (B) On the other hand, the replication of an eukaryotic chromosome begins simultaneously at several specific sites and proceeds in both the directions from each of these points. The arrows indicate the directions in which replication progresses.

of thin wool (Fig. 4.1B). As prophase progresses, chromosomes become increasingly shorter and thicker due to increased condensation. By mid-prophase, the two chromatids of each chromosome become visible (Fig. 4.1C). By the end of prophase, all the chromosomes become considerably shorter and thicker. (2) During prophase, nucleolus and nuclear membrane remain present. (3) The spindle apparatus is not yet organised.

Chromosome ***condensation,*** that is a decrease in length with an accompanied increase in thickness, is mainly due to coiling of chromosomes. In early prophase, the number of coils in each chromosome is very large but their diameter is relatively

small. There is a progressive increase in the diameter of coils so that their number per chromosome becomes correspondingly smaller. This leads to a progressive increase in the thickness and a resultant reduction in the length of chromosomes.

Metaphase
- Nucleolus disappears.
- Nuclear envelope disappears.
- Spindle apparatus present.
- Chromosomes organized at *equatorial plate, i.e.,* the plane in the middle of the two poles of spindle apparatus.
- Chromosomes shortest and thickest.

4.5.3. Metaphase

At the end of prophase, four important events take place: (1) disappearance of nucleolus, (2) breakdown of nuclear envelope and distribution of its components in the endoplasmic reticulum, (3) appearance of spindle apparatus and (4) arrangement of the chromosomes onto a single plane called equatorial plate.

Equatorial plate lies in the middle of the two poles of the spindle apparatus. But it may not necessarily be the central plane of the cell. Further, it is an imaginary plane and it does not represent any structural feature of the cell. At metaphase, *centromeres of all the chromosomes lie on the equatorial plate, while their arms may extend outside this plate.*

As a rule, the centromere of one of the two sister chromatids of a chromosome gets attached to a spindle fiber originating from one pole, while that of the other sister chromatid becomes attached to a spindle fiber emanating from the opposite pole. The shortening and lengthening of spindle fibers is primarily responsible for their movement to the equatorial plate.

The main features of metaphase are as follows: (1) absence of nucleolus, (2) disappearance of nuclear membrane, (3) arrangement of chromosomes on the equatorial plate, (4) the shortest and the thickest state of chromosomes, and (5) presence of spindle apparatus.

Anaphase
- Sister chromatids separate, beginning from centromere.
- Sister chromatids move to the opposite poles.
- Chromosomes shorter than at metaphase.
- Chromosome movement.
  - (*i*) due to shortening of chromosomal fibers of spindle apparatus.
  - (*ii*) requires energy from ATP.

4.5.4. Anaphase

During anaphase, the two sister chromatids of each chromosome separate and migrate towards the opposite poles of the cell. *Anaphase begins when the centromeres of chromosomes appear to divide longitudinally. The sister chromatids now begin to separate from each other. This stage ends with the reaching of the sister chromatids to the opposite poles.* Centromere is the first part of the chromosomes to begin to move towards the poles. The two arms of each chromatid seem to be dangling from the centromere towards the equatorial plate. Often *chromosomes become somewhat more condensed during anaphase as compared to metaphase.*

Chromosome movement during anaphase and just prior to metaphase is primarily due to spindle fibers. Spindle fibers originate at two points, which are located near the periphery of a cell and opposite each other. These points are known as ***poles***. In case of animal cells, each pole contains a pair of centrioles. But centrioles are not found in plants cells.

Some spindle fibers pass from one pole to the other; they are known as ***continuous fibers***. In contrast, ***chromosomal fibers*** pass from one pole to the centromere of a chromosome. Centromeres of sister chromatids, as a rule, become attached to the chromosomal fibers originating from opposite poles. The movement of sister chromatids during anaphase is primarily due to the shortening of chromosomal fibers. Chromosome movement during cell division requires energy, which is provided by ATP.

Telophase
- Chromosomes uncoil.
- They become wound into a loose ball.
- Nucleolus reappears.
- Nuclear envelope reappears.
- End of karyokinesis.

4.5.5. Telophase

Anaphase ends and telophase begins when sister chromatids of all the chromosomes of a cell reach the opposite poles. During telophase, the following events occur. (1) *The chromosomes uncoil so that they become very long and thin.* (2) *They also become wound into a loose ball of fine thread.* (3) *Nucleolus reappears, and* (4) *nuclear membrance is reorganised around each group of chromosomes.* (5) *At the end of telophase, phragmoplast appears at the equatorial plate of the cell.*

When nucleolus dissociates at the end of prophase, nucleolar materials are most likely distributed onto the chromosomes. During telophase, these materials separate from the chromosomes and contribute to the reorganisation of the new nucleolus. Similarly, the nuclear envelope dissociates into small elements, which become part of endoplasmic reticulum of the cell. During telophase, these elements reorganise around the two groups of chromosomes and fuse to produce nuclear envelope around them.

In terms of duration, prophase is the longest stage of the division phase of cell cycle. In comparison, anaphase is the shortest stage, while metaphase and telophase are considerably longer than anaphase. Often metaphase is somewhat longer than telophase. *Telophase marks the end of karyokinesis or nuclear division.*

4.6. CELL CYCLE

A meristematic cell cyclically passes through mitosis, G_1, S, G_2 and mitosis repeatedly. This phenomenon is termed as ***cell cycle*** and is represented as such (Fig. 4.4). RNA and protein syntheses are abundant during G_1 and G_2; they are depressed during S, prophase and upto early to mid metaphase; and absent during late metaphase, anaphase and upto early telophase.

In cell cycle, the division phase is the shortest, *e.g.*, 1 hr in *Vicia faba* (Table 4.1). The interphase has the longest duration, *e.g.*, 30 hr in *V. faba;* of these 30 hrs, G_1 and G_2 are 12 hr each, while S in only 6 hr. There is considerable variation in the

TABLE 4.1
Duration of different phases of cell cycle in certain organisms

	Duration (hour)			
	Interphase			Mitosis
Organism	G_1	S	G_2	
Vicia faba (root tip)	12	6	12	1
Mouse L cells (tissue culture)	12	6-8	3-4	1
Human He La cells (tissue culture)	12	10	3	1

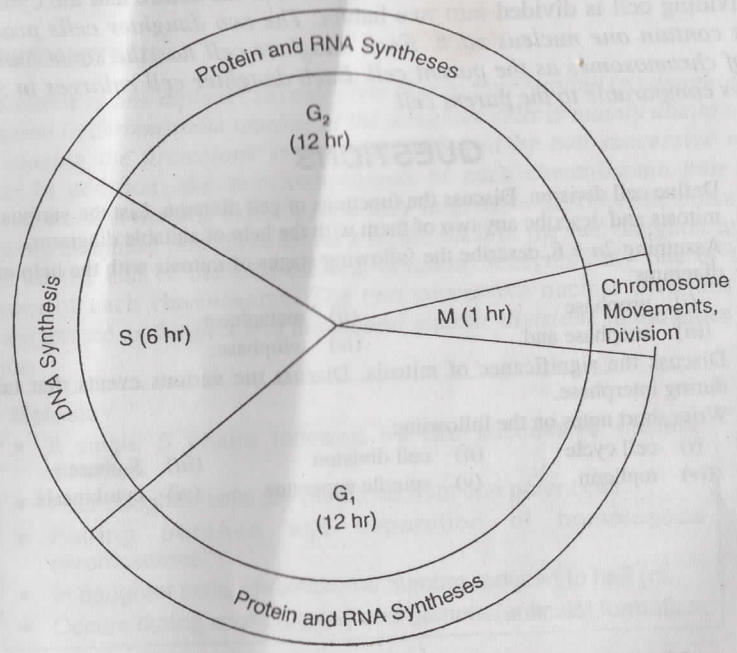


Fig. 4.4. Mitotic cell cycle in root-tips of *Vicia faba. M,* mitotic phase; G_1, first gap; S, DNA synthesis phase; and G_2, second gap. G_1, S and G_2 are the three substages of interphase. The time shown in the inner circle represents the approximate duration of the concerned stage.

duration of G_1. In some cells, G_1 may be very short or absent. The duration of G_2 is generally between 1–4 hr. But in some cases, G_2 may be upto 12 hr, and in plant species, like most fungi, the nondividing cells are in G_2 phase. The S phase generally occupies about 30–40% of the interphase time.

Cytokinesis
- At the end of telophase, phragmoplast appears at equatorial plate.
- Phragmoplast gives rise to cell plate.
- *In plants:* division of cytoplasm begins in center and progresses to periphery.
- *In animals*: cytoplasm division due to invagination of plasma lemma all round the cell .

4.7. CYTOKINESIS

Ordinarily, cytokinesis is complete by the end of telophase. At the equatorial plate, three different kinds of membranous elements organise into the phragmoplast. This gives rise to cell plate and, subsequently, to middle lamella.

In plants, the separation of cytoplasm begins in the centre of the cell. It gradually extends outward on each side in a plane perpendicular to the axis of the spindle. In animals, the division of cytoplasm begins with a simultaneous furrowing or invagination on all the sides in plasma lemma in the region of the equatorial plate. Finally, the invaginations of plasma lemma meet in the centre and the cytoplasm of each dividing cell is divided into two halves. *The two daughter cells produced by mitosis contain one nucleus each. Each daughter cell has the same number and kinds of chromosomes as the parent cell. Each daughter cell enlarges in size till it becomes comparable to the parent cell.*

QUESTIONS

1. Define cell division. Discuss the functions of cell division. List the various stages of mitosis and describe any two of them with the help of suitable diagrams.
2. Assuming $2n = 6$, describe the following stages of mitosis with the help of suitable diagrams:
   (*i*) prophase (*ii*) metaphase
   (*iii*) anaphase and (*iv*) telophase.
3. Discuss the significance of mitosis. Discuss the various events that take place during interphase.
4. Write short notes on the following:
   (*i*) cell cycle (*ii*) cell division (*iii*) *S*-phase
   (*iv*) replicon (*v*) spindle apparatus (*vi*) cytokinesis.

5

Cell Division II. Meiosis

5.1. INTRODUCTION

Meiotic division** of one diploid* (2n) *cell gives rise to four haploid* (n) *daughter cells. The reduction in chromosome number of the daughter cells is mainly due to a single S phase (during the premeiotic interphase) followed by two successive nuclear divisions.* In addition, the two homologues of each chromosome pair during prophase of the first meiotic division, and they migrate to the opposite poles during the following anaphase (anaphase I). As a result, each of the two daughter nuclei so produced (at the end of the first meiotic division) receives only one of the two homologues of each chromosome. The two successive nuclear divisions during meiosis are termed as the ***first and the ***second meiotic divisions*** in the order of their occurrence.

Meiosis
- A single *S* phase followed by two successive nuclear divisions.
- Four daughter cells are produced from one parent cell..
- Pairing between and separation of homologous chromosomes.
- In daughter cells, chromosome number reduced to half (*n*).
- Occurs during spore (plants) and gamete (animals) formation.

5.2. OCCURRENCE

In general, *meiosis takes place during gamete (animals)/spore (plants) formation.* Therefore, *it is confined to reproductive cells.* In plants, meiosis occurs in anthers and in ovules present in the ovaries. In case of animals, it takes place in testes (males) and ovaries (females).

Meiosis: Functions
- Production of haploid (*n*) gametes.
- Release of genetic variation by:
  1. Segregation.
  2. Independent assortment.
  3. Crossing over.

5.3. SIGNIFICANCE OF MEIOSIS

As a consequence of meiosis, gametes contain only half (*n*) of the somatic chromosome number (2*n*). Therefore, union between one male and one female

gamete during fertilization restores the chromosome number to the diploid ($2n$) state. Thus the chromosome number of a species remains constant from one generation to the next generation produced by sexual reproduction. The various functions of meiosis may be summarised as follows.

(1) ***Production of haploid (n) gametes*** so that fertilization restores the normal somatic ($2n$) chromosome number.

(2) ***Production of genetic variation*** through segregation, independent assortment and recombination.

(3) ***Segregation of the two alleles of each gene.*** This takes place due to pairing between the two homologues of each chromosome and their separation at the first anaphase.

(4) ***Independent segregation of the alleles of different genes*** located in separate chromosomes. This takes place due to independent orientation of different bivalents at the first metaphase.

(5) ***Recombination between linked genes*** due to crossing over during pachytene stage.

5.4. STAGES OF MEIOSIS

In contrast to mitosis, meiosis is spread over a considerably long duration. *During meiotic division, the nucleus of each cell undergoes two successive divisions, which are referred to as the first and the second meiotic divisions.*

Each of the two meiotic divisions is divided into prophase, metaphase, anaphase, and telophase. The word *first* is added to the names of the various stages of the first meiotic division to yield first prophase (prophase I), first metaphase (MI), first anaphase (AI) and first telophase (telophase I). Similarly, the word *second* is added to the names of the different stages of the second meiotic division to give second prophase (prophase II), second metaphase (MII), second anaphase (AII) and second telophase (telophase II).

Premeiotic Interphase

- The interphase just before a cell enters meiosis.
- Biochemicals produced during *G2* are essential for the onset of meiosis.
- 0.3% of DNA does not replicate; this may be involved in chromosome pairing.

5.5. PREMEIOTIC INTERPHASE

The interphase stage just prior to the entry of a cell into meiosis in known as ***pre-meiotic interphase.*** *The events taking place during pre-meiotic interphase are crucial for the entry of cells into meiosis.* During the *S* phase of pre-meiotic interphase, chromosome replication takes place. But *approximately 0.3% of the total DNA present in the nucleus does not replicate during this S phase. This DNA replicates during the zygotene substage of prophase I.*

The entry of cells into meiosis depends on some substances synthesized during G_2 *substage of pre-meiotic interphase.* During meiosis, the *S* phase occurs only once during the pre-meiotic interphase. *Thus, during meiosis chromosomes replicate only once, while the nuclei divide twice.*

5.6. FIRST MEIOTIC DIVISION

The most significant events of the first meiotic division are as follows: (1) pairing between homologous chromosomes, (2) crossing over between them during prophase I, and (3) separation of homologous chromosomes and their migration to the opposite poles of the cell during AI. As a result, ***the two daughter nuclei produced by this division receive only half of the chromosomes present in somatic cells.*** For this reason, the first division is often referred to as the ***reduction division.*** There are four stages of this division: (1) prophase I, (2) metaphase I (MI), (3) anaphase I (AI), and (4) telophase I.

First Prophase

- Longest in duration:
  - (*i*) *Leptotene*: Chromosomes become visible; RNA and protein synthesis.
  - (*ii*) *Zygotene*: Synapsis; the remaining 0.3% DNA replicates; synaptonemal complex develops.
  - (*iii*) *Pachytene*: Chromosomes become shorter, thicker, and are visible as *n* bivalents; crossing over takes place; Ca. 0.3% DNA synthesis.
  - (*iv*) *Diplotene*: Homologues begin to separate; chiasma terminalization.
  - (*v*) *Diakinesis*: Chiasma terminalisation is complete when it begins; *n* bivalents.

5.6.1. First Prophase

This stage is the longest in duration. In female animals, it may last for several days to several years. Prophase I is divided into five substages: (1) leptotene, (2) zygotene, (3) pachytene, (4) diplotene, and (5) diakinesis.

5.6.1.1. Leptotene

During leptotene, (1) there is a marked increase in the nuclear volume. (2) *There is chromosome condensation so that they become visible as fine threads.* All the chromosomes in a nucleus appear wound round each other like a loose ball of wool (Fig. 5.1A) (3) *There is RNA synthesis as a result of which the volume of nucleolus increases appreciably.* (4) Proteins required for chromosome condensation are synthesized.

5.6.1.2. Zygotene

This stage begins with the initiation of pairing between homologous chromosomes, and it ends when the pairing is completed (Fig. 5.1B). The main events occurring during this stage may be summarized as follows: (1) *pairing between homologous chromosomes, (2) completion of replication of the remaining 0.3% DNA,* (3) synthesis of a specific nuclear protein, (4) *development of the synaptonemal complex,* and (5) progressive condensation of chromosomes.

Pairing between homologous chromosomes is extremely precise. Only the homologous segments of the two chromosomes pair with each other. In polyploid strains (or nuclei) containing more than two homologues of each chromosome, *only two homologous chromosomes pair in any given segment of a chromosome.* In no circumstance more than two homologues can pair in the same segment. Pairing

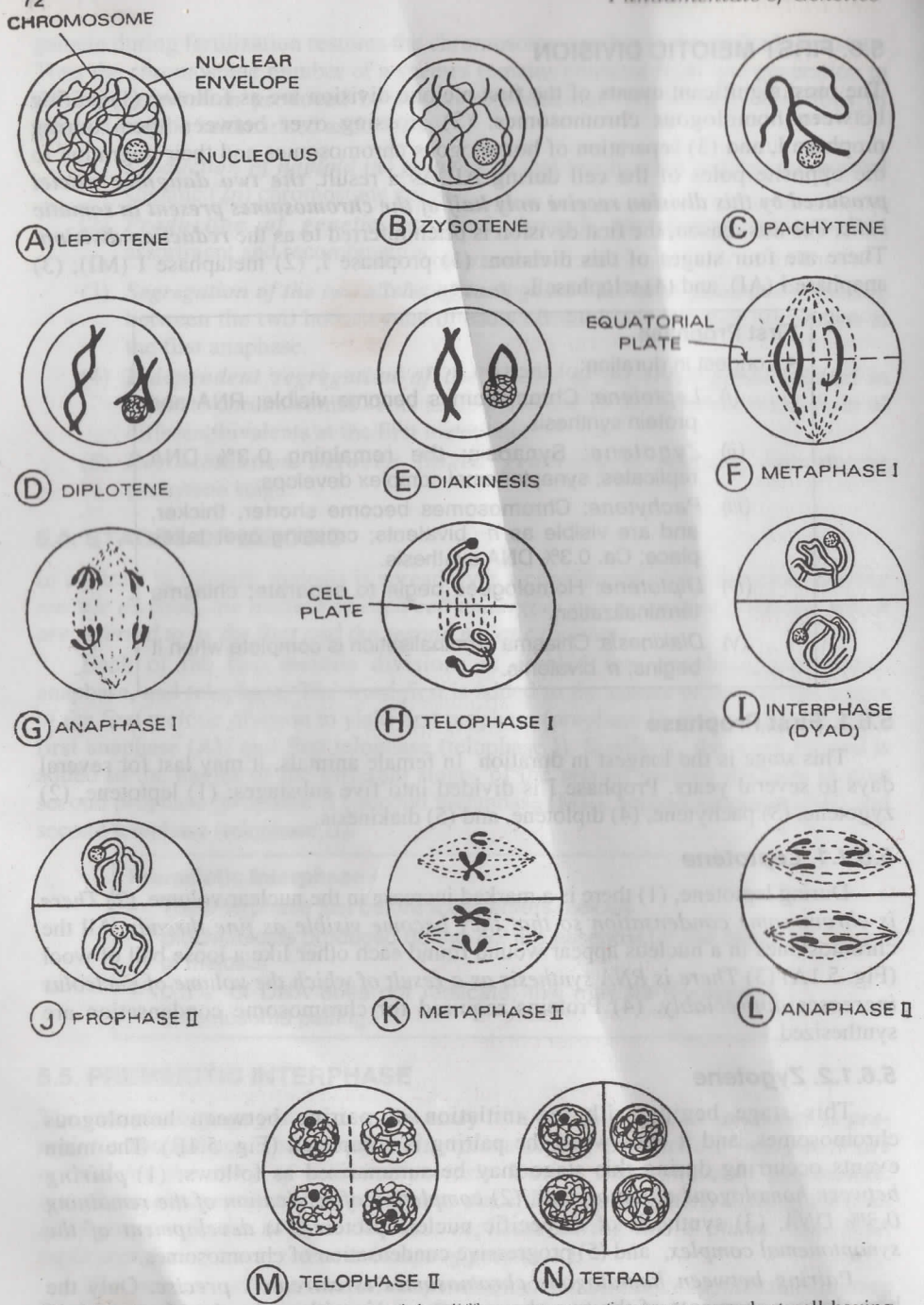


Fig. 5.1. A schematic representation of the different stages of meiosis in a plant cell having $2n = 4$ chromosomes. In this species, telophase I and the following interphase are shown as is the case in maize. In some species, *e.g., Trillium,* telophase I and the following interphase are absent, and cells enter prophase II directly from anaphase I.

between homologous chromosomes is often referred to as ***synapsis***. Synapsis (1) may begin at both the ends of a homologous pair and proceed towards their centre; (2) it may begin at the centromere and progress towards the telomeres; or (3) it may begin simultaneously at several places.

The reasons for synapis are unknown. The ***precocity theory*** of synapsis, proposed by Darlington, seemed to explain the phenomenon quite satisfactorily. But it is unacceptable in the face of compelling experimental evidence against it. According to this theory, each chromosome consists of a single chromatid during leptolene and zygotene. It has been suggested that chromosomes have a strong tendency for a paired structure. As a result, homologous chromosomes are attracted towards each other to form pairs. They were postulated to replicate during pachytene. As a result, each of the two homologues would now be composed of a pair of chromatids. Therefore, the need for being paired would be satisfied. As a result, they would now begin to repel each other and move away from each other during diplotene and diakinesis. Clearly, *this theory requires that the nuclei enter into meiosis before their chromosomes replicate. This also makes it necessary that the chromosome replication should take place during pachytene. However, it is certain that chromosomes replicate during the premeiotic interphase.* Consequently, each chromosome is already double-stranded during the leptontene itself. Therefore, synapsis cannot be due to the attraction between unreplicated chromosomes, and ***the precocity theory is unacceptable.***

A small part of DNA (approx. 0.3%) replicates during zygotene. This is essential for chromosome pairing. It was postulated by Stern that several segments of chromosomal DNA remain single-stranded due to the incomplete DNA replication during premeiotic interphase. These single-stranded regions are proposed to play a significant role in chromosome pairing.

Synaptonemal Complex
- Appears in zygotene between homologous chromosomes.
- Composed of protein.
- *3 Elements*:
  - (*i*) *Central*: one; dense element.
  - (*ii*) *Laterral*: two elements: dense, associated with the homologous cromosomes.
  - (*iii*) *Transverse*: two elements, less dense; join the central element to lateral elements.
- Role in crossing over not definite.

1. ***Synaptonemal Complex.*** As a rule, chromosome pairing occurs in two stages. In the ***first stage***, homologous chromosomes move close to each other and stay about 1000 Å away from each other. In the ***second phase,*** a special structure, called synaptonemal complex, develops in this 1000 Å region.

Synaptonemal complex is composed of protein. It consists of three distinct elements: (1) a relatively dense ***central element***, (2) two comparatively dense ***lateral elements***, each of which is adhered to one of the two homologues through its entire length, and (3) two relatively less dense ***transverse elements*** (Fig. 5.2).

The function of synaptonemal complex is uncertain and controversial. It has been suggested that synaptonemal complex functions in and is essential for crossing over. In male *Drosophila,* synaptonemal complex is absent, and there is no recombination between linked genes either. This fact tends to support the above

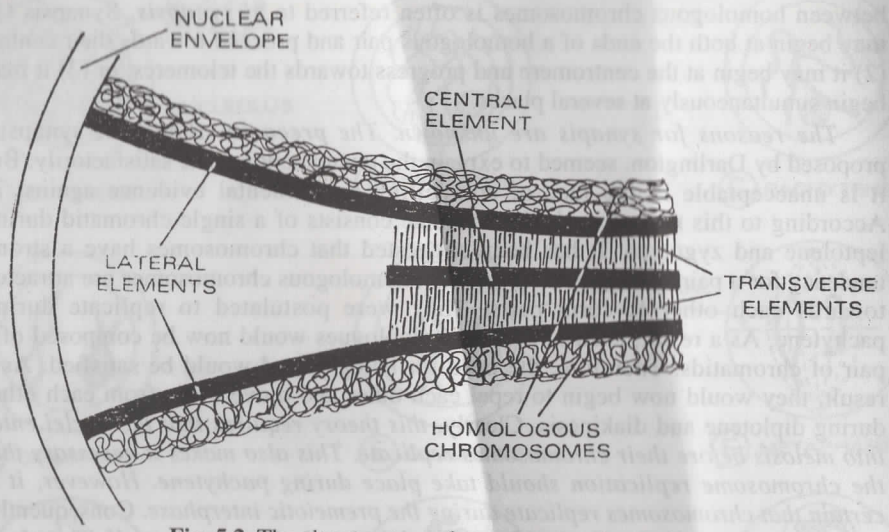


Fig. 5.2. The ultrastructure of a synaptonemal complex.

conclusion. But in some cases, synaptonemal complex is found between nonhomologous chromosomes as well; this seriously questions its role in crossing over.

5.6.1.3. Pachytene

Pachytene begins when synapsis comes to an end. This stage ends when the homologous chromosomes begin to move away from each other. The main events that occur during this stage are as follows. (1) There is further condensation of chromosomes so that chromosome pairs become shorter ad thicker. (2) Since the homologous chromosomes have already paired, *the number of observable chromosomes is only half of the somatic chromosome umber, that is, it is only* n (Fig. 5.1C). (3) In most species, chromosome pairs are easily recognisable during this stage, *e.g.,* in maize. Before pachytene, the bivalents are too long to be identified individually. (4) The nucleolus is distinct and quite large. It is associated with the nucleolus organiser region (NOR) of the sat-chromosome pair. (5) *Crossing over between homologous chromosomes takes place during this stage.* (6) *A small amount* (*Ca.* 0.3%) *of DNA synthesis takes place during pachytene. This DNA synthesis is believed to play a crucial role in crossing over.*

The structure produced by pairing between two homologous chromosomes is known as ***bivalent.*** Thus each bivalent is composed of two chromosomes. In turn, each chromosome consists of two chromatids. (Remember, chromosome replication has already taken place during the premeiotic prophase.) *Thus each bivalent contains four chromatids (or strands). Therefore, the bivalents are said to be in the four-strand stage or tetrad stage.*

5.6.1.4. Diplotene

The main events taking place during diplotene are as follows. (1) *The homologous chromosomes making up each bivalent begin to move away from each other.* This marks the beginning of this stage. Thus the two chromosomes constituting each bivalent become clearly visible. (2) The two homologues of each

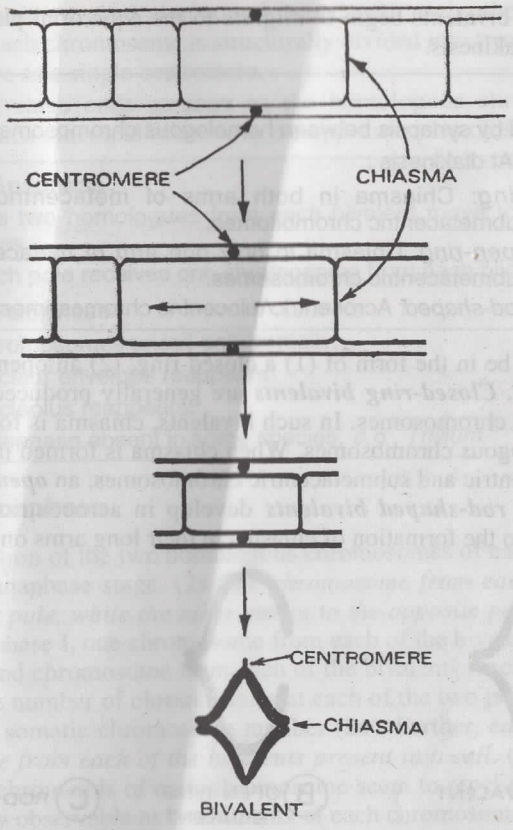


Fig. 5.3. A schematic representation of chiasma terminalization. As the homologous chromosomes move away from each other, the chiasma are pushed away from near the centromere towards the telomeres (indicated by arrows). In the end, only the two telometres of the homologues remain in contact with each other.

bivalent appear to be attached with each other at one or more points. These attachments are known as ***chiasmata*** (sing., chiasma) (Fig. 5.1D). (3) As diplotene progresses, chiasmata slowly move toward the ends of homologous chromosomes. This movement is referred to as ***chiasma terminalization,*** *i.e.,* movement of chiasma towards terminal positions in the chromosomes (Fig. 5.3). *Chiasma terminalization occurs mainly due to the movement of homologous chromosomes away from each other.* (4) There is further condensation of chromosomes so that they become progressively shorter and thicker.

5.6.1.5. Diakinesis

The end of chiasma terminalization marks the beginning of this stage. The main events of diakinesis are as follows: (1) Chiasma terminalization is completed just before this stage. *The two homologous chromosomes of each bivalent are now attached at or close to the telomeres only* (Fig. 5.1E). (2) Chromosomes become shorter and thicker due to further condensation. (3) *Bivalents move away from each other and spread towards the peripheri of cells.* (4) Nucleolus and nuclear envelope disappear towards the end of diakinesis. (5) The spindle apparatus becomes

organised. (6) The bivalents begin to migrate to the equatorial plate of cells. This marks the end of diakinesis.

Bivalent
- Formed by synapsis between homologous chromosomes.
- *Types*: At diakinesis :
  - (*i*) *Ring*: Chiasma in both arms of metacentric and submetacentric chromosomes.
  - (*ii*) *Open-ring*: Chiasma in only one arm of metacentric/ submetacentric chromosomes.
  - (*iii*) *Rod-shaped*: Acrocentric/tolocentric chromosomes.

Bivalents may be in the form of (1) a closed-ring, (2) an open-ring or (3) they may be rod-shaped. ***Closed-ring bivalents*** are generally produced by metacentric and submetacentric chromosomes. In such bivalents, chiasma is formed in both the arms of the homologous chromosomes. When chiasma is formed in only one of the two arms of metacentric and submetacentric chromosomes, an ***open-ring bivalent*** is produced. Finally, ***rod-shaped bivalents*** develop in acrocentric and telocentric chromosomes due to the formation of chiasma in their long arms only (Fig. 5.4).

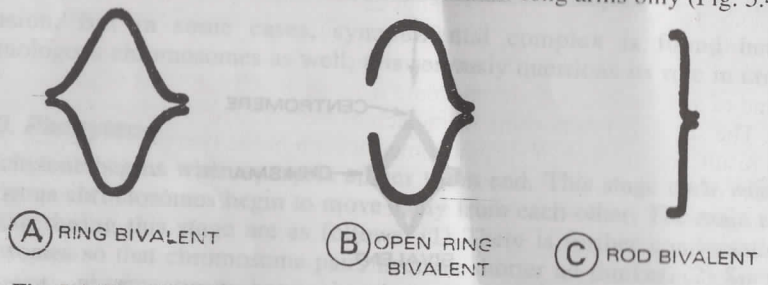


Fig. 5.4. The different types of bivalents seen during metaphase I of meiosis.

First Metaphase
- Nuclear envelope disappears.
- Nucleolus disappears.
- Spindle apparatus becomes organized.
- Bivalents become arranged at equational plate.

5.6.2. First Metaphase

Just prior to the beginning of metaphase, (1) *nucleolus disappears*, (2) *nuclear envelope disintegrates* and its elements become a part of ER, (3) *spindle apparatus becomes organised*, and (4) all the *bivalents within a cell migrate to the equatorial plate*. Centromeres of the two homologous chromosomes of each bivalent become attached to chromosomal fibers originating from the opposite poles. Movement of bivalents on to the equatorial plate is primarily due to the shortening/lengthening of these chromosomal fibers.

During metaphase, (1) bivalents are arranged at the metaphase plate. (2) The centromeres of the two homologues of each bivalent lie on either side of the equatorial plate; (3) the part of the bivalents lying on the equatorial plate is the central part of each bivalent, which consists of the telomeres of the two homologous

chromosomes now attached almost end-to-end with each other (Fig. 5.1). (4) The centromere of each chromosome is structurally divided into two, but functionally the two parts behave as a single centromere.

Metaphase stage ends as soon as the homologous chromosomes begin to separate from each other and to move to the opposite poles of the cell.

First Anaphase
- The two homologues from each bivalent move to opposite poles.
- Each pole receives one chromosome from each bivalent.

First Telophase
- Chromosomes uncoil only partially.
- Nuclear envelope reappears.
- Nucleolus reappears.
- Telophase absent in some species, *e.g.*, *Trillium*.

5.6.3. First Anaphase

(1) Separation of the two homologous chromosomes of each bivalent marks the beginning of anaphase stage. (2) *One chromosome from each bivalent begins to migrate to one pole, while the other moves to the opposite pole* (Fig. 5.1G). (3) At the end of anaphase I, one chromosome from each of the bivalents gathers about one pole. The second chromosome from each of the bivalents reaches the opposite pole. As a result, the number of chromosomes at each of the two poles of a cell is exactly half (n) of the somatic chromosome number ($2n$). Further, *each pole receives only one homologue from each of the bivalents present in a cell*. (4) During this phase, the two sister chromatids of each chromosome seem to repel each other. Therefore, they are clearly observable as constituents of each chromosome. *This stage comes to an end when the homologous chromosomes reach the opposite poles.*

5.6.4. First Telophase

During this stage, (1) the *chromosomes uncoil only partially*, while a complete uncoiling takes place during mitotic telophase. (2) *Nuclear envelope becomes organised* around the two groups of chromosomes. (3) *Nucleolus also reappears* (Fig. 5.1H).

In some species, *e.g.*, *Trillium*, cells enter directly into the second prophase from the first anaphase. In these species, cells do not enter into the first telophase and the following interphase.

Cytokinesis
- In maize etc., cytoplasm divides into two.
- The two celled structure called dyad.
- In *Trillium* etc., occurs after second division only.

Interphase
- Absent in some species, *e.g.*, *Trillium*.
- NO DNA SYNTHESIS.
- No further uncoiling of chromosomes.

5.6.5. Cytokinesis

In many species, *e.g.*, maize, the cytoplasm of each cell divides into two halves by the end of first telophase. In each of the two daughter cells thus obtained, a single haploid nucleus is present. The two daughter cells do not separate, but they stay together; this two-celled structure is known as a ***dyad***. In some species, on the other hand (*i.e., Trillium*), cytokinesis occurs only at the end of the second meiotic division.

5.7. INTERPHASE

In many species, the interphase after first meiotic division is absent. In other species, where it occurs, it is very short in duration. *It is important to note that* ***there is no DNA synthesis during this interphase.*** In addition, chromosomes do not uncoil any further than they already have at the end of telophase I (Fig. 5.1I).

5.8. SECOND MEIOTIC DIVISION

During the second division of meiosis, the two haploid nuclei of each dyad divide synchronously. Ordinarily, the plane of second division is perpendicular to that of the first meiotic division. But during megasporogenesis (in flowering plants) and ascospore development (in Ascomycetes), the two divisions occur in the same plane.

During the second division, *the two sister chromatids of each chromosome separate and migrate to the opposite poles.* As a result, the number of chromosomes in each of the haploid nuclei remains the same (*i.e.,* haploid) at the end of this division. Therefore, the second division is often referred to as ***equational division.*** There are a number of similarities between this division and mitosis; hence, it is sometimes called *meiotic mitosis*. This, however, is not desirable as it may create confusion without clarifying anything.

The second meiotic division is also divided into four stages as follows: (1) prophase II, (2) memtaphase II (MII), (3) anaphase II (AII) and (4) telophase II.

Second Prophase
- The two nuclei of a dyad divide synchronously.
- *n* chromosomes become visible .
- The two chromatids of each chromosome are clearly visible.

5.8.1. Second Prophase

This stage is quite similar to the prophase of mitosis. But there are several important differences between the two. (1) *During this stage, there is no relational coiling between sister chromatids.* (2) *The two sister chromatids of each chromosome are clearly visible* (Fig. 5.1J). (3) The chromosomes are relatively much more condensed so that they appear shorter and thicker than in mitosis of the same species. (4) There is further condensation of chromosomes so that they become thicker and shorter. (5) *The chromosome number is only* n, *while in mitosis it is* 2n.

Second Metaphase
- Nuclear envelope disappears.
- Nucleolus disappears.
- Spindle apparatus becomes organized.
- Chromosomes are arranged on equatorial plate.

At the end of this stage, nucleolus and nuclear envelope disappear, and spindle apparatus is organised. Following this, chromosomes migrate to the equatorial plate.

5.8.2. Second Metaphase

This stage begins when the *chromosomes become arranged on the equatorial plate.* During this stage, (1) *nucleolus and* (2) *nuclear envelope are absent,* (3) *spindle apparatus is present*, and (4) *centromeres of all the chromosomes are arranged in a single plane at the equatorial plate.* (5) The two sister chromatids of each chromosome are distinctly separated from each other, but they remain attached at the centromere (Fig. 5.1K). (6) Chromosomes become more condensed, thicker and shorter. (7) This stage is relatively quite short in duration.

5.8.3. Second Anaphase

Anaphase II begins when (1) centromeres of chromosomes observably divide longitudinally, and *the two sister chromatids of each chromosome begin to separate and move away to the apposite poles* (Fig. 5.1L). (2) This stage comes to an end when the sister chromatids of chromosomes reach the opposite poles.

Second Anaphase
- Sister chromatids of each chromosome move to opposite poles.
- Each pole receives one sister chromatid from each chromosome.

Second Telophase
- Chromatids uncoil.
- Nuclear envelope reappears.
- Nucleolus reappears.

Cytokinesis
- In many species, occurs after first and second telophases.
- In some species, occurs only after second telophase.
- Four haploid cells produced from each cell entering meiosis.

In electron micrographs, the centromere of each chromosome is structurally divided into two during the first meiotic division itself. However, the centromeres of the two sister chromatids of each chromosome become functional and appear divided visibly (under light microscope) only during A II.

5.8.4. Second Telophase

Second telophase begins when sister chromatids of chromosomes reach the opposite poles. During this stage, (1) *the chromatids uncoil* so that they assume the appearance of a loose ball of thread, (2) *nuclear envelope is reorganised* from the elements of ER, and (3) *nucleolus reappears* (Fig. 5.1M).

9. CYTOKINESIS

At the end of telophase II, the cytoplasm of each of the two cells of a dyad divides into two parts. As a result, *one parent cell produces four haploid daughter cells after*

the completion of meiotic division. The four daughter cells produced from a parent cell are present together, and are together known as ***tetrad*** (Fig. 5.1N). In some species, dyads are not produced. As a result, at the end of the second telophase, four haploid nuclei are present in a single cell. Cytokinesis occurs only at the end of telophase II. Therefore, each cell is now divided into four daughter cells each of which contains a single haploid nucleus.

The four haploid cells produced by meiotic division of a single cell may differentiate into gametes (animals) or spores (plants). The spores ordinarily give rise to the gametophytic generation, i.e., the generation, which produces gametes through mitosis. This generation may be independent (lower plants) or totally dependent (higher plants) on the diploid sporophytic stage, that is, the generation that gives rise to the spores through meiosis. *In plants, gametes are generally produced through mitosis in the gametophyte.*

5.10. DIFFERENCES BETWEEN MITOSIS AND MEIOSIS

Meiosis is undoubtedly derived from an extensive modification of the mitotic division. But both the divisions, as they occur today, differ with respect to a number of important features (Table 5.1).

TABLE 5.1
Differences between mitosis and meiosis

Characteristic	*Mitosis*	*Meiosis*
Location	*Somatic and reproductive tissues*	*Reproductive tissues only.*
Shape of cell	Generally, rectangular.	Generally, spherical.
Daughter cells	One cell produces two daughter cells.	On completion of meiosis, four daughter cells are produced from one cell.
	Chromosome number of daughter cells is the same as that of the parent cell.	*Chromosome number of daughter cells is half of that of the parent cell.*
Genetic variation	*Maintains genetic uniformity among daughter cells.*	*Generates genetic variability due to segregation and recombination.*
S *phase*	*Occurs in every interphase following mitosis.*	*Occurs only in the premeiotic interphase; there is no* S *phase in the interphase following the first division of meiosis.*
	The total DNA present in a nucleus is replicated during S *phase.*	*About 0.3% of the DNA is not replicated during premeiotic* S *phase.*
Number of nuclear divisions	*Each nucleus divides once after every* S *phase.*	*After the* S *phase, nucleus undergoes two successive divisions.*

Prophase		***First prophase***
	Homologous chromosome do not pair; 2*n* chromosome number can be seen.	*Homologous chromosomes pair.* As a result, only *n* chromosomes can be seen.
	There is no crossing over between homologous chromosomes, and chiasma cannot be seen.	*There is crossing over* between homologous chromosomes, and chiasma are observable.
	No DNA synthesis.	*About* 0.3% *of the DNA is synthesized during zygotene and pachytene.*
		Second Prophase
	Chromosomes are 2*n in number.*	Chromosome number is only *n.*
	Sister chromatids show relational coiling.	Relational coiling between sister chromatics is absent.
	In early prophase, chromosomes are very long; later they become smaller due to progressive condensation.	Chromosomes are highly condensed and small from the very beginning.
	It is quite long in duration.	*It is of relatively short duration.*
Metaphase		***First Metaphase***
	Each chromosome is unpaired.	*Homologous chromosomes are paired to form bivalents.*
	2n *chromosomes are clearly seen.*	*The number of bivalents (chromosome pairs) is* n.
	Centromeres of all the chromosomes lie on the equatorial plate.	Centromeres of chromosomes lie on the either side of equatorial plate.
	Each chromosome has two chromatids.	Each bivalent has four chromatids.
		Second Metaphase
	The number of chromosomes is 2n.	*The chromosome number is* n.
	The two sister chromatids of a chromosome do not exhibit repulsion.	Repulsion between sister chromatids is clearly seen from prophase itself.
	Chromosomes are relatively longer.	Chromosomes are much shorter.
Anaphase		***First Anaphase***
	The two sister chromatids of each chromosome move to the opposite poles.	*Homologous chromosomes from each bivalent move to the opposite poles.*
	Centromere of each chromosome divides longitudinally.	*Centromeres do not divide.*

		Second Anaphase
	Chromosome number is 2n.	*Chromosome number is* n.
	Generally, a single nucleus divides in a cell.	In all the stages of second meiotic division, two nuclei divide in one parent cell.
		First Telophase
Telophase and cytokinesis	Chromosomes almost completely uncoil.	Chromosomes do not uncoil completely.
	In all the species, cytokinesis takes place.	In some species, cytokinesis does not take place.
	Cytokinesis involves the formation of middle lamella and primary cell wall.	Cytokinesis consists of membrane formation, and middle lamella *is not* deposited
	Telophase and interphase are found in all species.	*In some species, first telophase and the following interphase are absent.*
		Second Telophase
	Each cell produces two daughter cells.	*Each cell gives rise to four haploid cells; the resulting structure is called* ***tetrad****.*
	Daughter cells are 2n.	*Daughter cells are* n.
	Daughter cells generally differentiate as body cells.	*Daughter cells generally differentiate into spores/gametes.*

QUESTIONS

1. List the various stages and substages of meiosis. Describe the different substages of prophase I giving suitable diagrams. Discuss the functions of meiosis.
2. Assuming $2n = 6$, briefly describe the following stages of meiosis with the help of suitable diagrams:
   (*i*) zygotene, (*ii*) pachytene, (*iii*) diplotene,
   (*iv*) diakinesis, (*v*) metaphase I, (*vi*) anaphase I
   (*vii*) anaphase II, and (*viii*) metaphase II.
3. Write short notes on the following:
   (*i*) importance of meiosis, (*ii*) premeiotic interphase,
   (*iii*) chromosome pairing, (*iv*) chiasma terminalization,
   (*v*) synaptonemal complex, (*vi*) diakinesis, and
   (*vii*) prophase I.
4. Differentiate between the following:
   (*i*) zygotene and pachytene, (*ii*) diplotene and diakinesis,
   (*iii*) mitotic metaphase and metaphase II, (*iv*) metaphase I and metaphase II,
   (*v*) anaphase I and anaphase II, (*vi*) mitotic anaphase and anaphase II,
   (*vii*) mitotic prophase and prophase II.
5. A plant has 20 chromosomes in its somatic cells. Give the chromosome number in the following cells of this plant:
   (*i*) root-tip cell, (*ii*) endosperm cell, (*iii*) zygote,
   (*iv*) leaf mesophyll cell, (*v*) pollen grain, (*vi*) megaspore mother cell,

   (*vii*) egg cell, (*viii*) secondary nucleus, (*ix*) pollen mother cell,
   (*x*) sperm, (*xi*) synergid, (*xii*) epidermal cell,
   (*xiii*) shoot-tip cell, (*xiv*) cotyledon cell.
6. Name the type of cell division (mitosis or meiosis), which leads to the production of the following:
   (*i*) pollen grain from PMC, (*ii*) embryo sac from megaspore,
   (*iii*) embryo from zygote, (*iv*) leaf from shoot apical meristerm,
   (*v*) megaspore from megaspore mother cell,
   (*vi*) male gametophyte from pollen grain, (*vii*) fruit from ovary,
   (*viii*) endosperm from primary endosperm nucleus.
7. During which stage of cell division, the following events take place ?
   (*i*) chromosome pairing or synapsis,
   (*ii*) DNA replication,
   (*iii*) movement of chromosomes to the equatorial plate,
   (*iv*) movement of chromosomes/chromatids to the poles,
   (*v*) organisation of the spindle fiber,
   (*vi*) crossing over,
   (*vii*) chiasma terminalization,
   (*viii*) separation of sister chromatids during meiosis,
   (*ix*) reorganisation of nuclear envelope,
   (*x*) disappearance of nucleolus/nuclear envelope,
   (*xi*) separation of homologous chromosomes during meiosis.

6

Regulation of Cell Cycle

6.1. INTRODUCTION

Meristematic cells cyclically pass through mitotic (***M phase***), G_1, S and G_2 phases of cell cycle (Fig. 4·3). A cell just produced by cell division is in G_1 phase. There is almost a continuous increase in protein and RNA contents of the cells during interphase leading to their doubling by the end of G_2 phase. The DNA content of cells, however, doubles during the S phase (Fig. 6·1). Thus a cell achieves the following two features before it begins to divide again.

1. *Every DNA sequence present in the cell nucleus is replicated once and only once.*
2. Cell mass increases to about twice as much as it was at the begining of G_1 phase. This is necessary to prevent a gradual reduction in cell mass with every cell division.

A cell must ensure the following before entering cell division:
- Almost doubling of cell mass.
- Replication of every regment of DNA once and only one.
- Absence of DNA damage, especially double strand breaks.

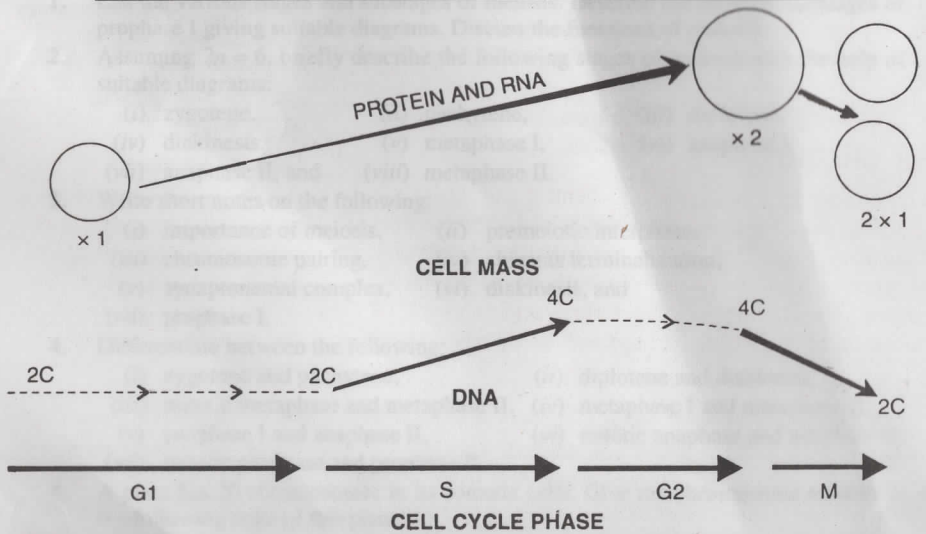


Fig. 6.1. A diagrammatic summary of the changes in cell mass and DNA content during somatic cell cycle.

The above two events are essential for the cell to survive the consequences of continued cell division. Therefore, the cell must regulate its cell cycle in relation to these two features.

6.2. DOMINO AND CLOCK THEORIES

The regulation of cell cycle has been explained on the basis of two theories called (1) domino theory and (2) clock theory. According to ***domino theory***, cell division is comparable to biosynthetic pathway in which a given reaction can occur only when the previous reaction in the pathway is successfully completed (Fig. 13·9). In this scheme, specific regulators that either activate or inhibit specific stages of cell division would regulate the progress of cell cycle. This theory was developed from analysis of *cdc* (***cell division cycle***) mutants of yeast, and the concept of checkpoints (Section 4) is integral to it.

The ***clock theory*** evolved from studies on control of mitosis in eggs and young embryos of sea urchin and *Xenopus* (a toad). According to this theory, cells oscillate between interphase and *M* phase in response to a timer or oscillator present in them. As a result, inhibition of a previous stage of cell cycle (say, of *S* phase) will not prevent the onset of the next stage (say, in this case, *M* phase). This concept adequately explains the findings from these systems, and seems to be distinct from the concept of domino theory. But *we do know now that similar regulatory molecules are involved both in yeast* cdc *mutants and in eggs and embryos of* Xenopus *and sea urchin.*

Checkpoints
- A point in cell cycle at which a cell decides whether or not to proceed with cell cycle :
  1. *START* (in yeast) or *restriction point* (in animal cells): Cells decide whether or not to enter *S* phase; major checkpoint in diploid organisms; early in *G1* phase.
  2. *Initiation of M phase*: Toward the end of *G2* phase; cells evaluate cell mass, DNA replication and damage; decision to enter M phase; major control point in haploid organisms.
- Checkpoints regulated by specific protein kinases.

6.3. CHECKPOINTS

The concept of checkpoints is central to regulation of cell cycle. A ***checkpoint*** is a specific point in cell cycle at which cells evaluate whether all the conditions necessary for cell division are fulfilled. If the conditions are fulfilled, cell cycle progresses past the checkpoint, otherwise it is blocked at the checkpoint. There are two major checkpoints in eukaryotes: (1) *START* or restriction point and (2) initiation of *M* (mitosis) phase.

6.3.1. Start

START is a point early in G_1 phase when cells become committed to enter cell division; this point in cell cycle is called ***START*** in yeast and ***restriction point*** in animal cells. At this checkpoint, a cell evaluates its internal as well as external environments and assesses if it is in a position to enter *S* phase. If it finds the conditions to be suitable, it decides to enter *S* phase; in fact, S phase begins some times after this decision is made. *In cultured animal cells, and in almost all diploid organisms, this is the main checkpoint. Therefore, a cell that enters S phase will*

normally enter M phase as well. This checkpoint is regulated by a protein kinase that is similar to M phase kinase (Section 6.6).

6.3.2. Initiation of M Phase

The second checkpoint is located towards the end of G_2 phase. At this checkpoint, a cell evaluates doubling of its cell mass, the completeness of DNA replication and absence of DNA damage. In all eukaryotes, completion of DNA replication is a prerequisite for the onset of *M* phase. In addition, the cell mass must increase to almost double and there should be no DNA damage. Each of these features are checked by specific proteins encoded by specific genes (Table 6·1). This checkpoint is the main control point in haploid organisms; it is regulated by a protein kinase called ***M phase kinase.***

TABLE 6.1

The various checkpoints and the nature of controls in regulation of cell division cycle. Most of the information is based on studies in yeast

Event	*Checkpoint*	*Stage of cell cycle*	*Mechanism*	*Agent*	*Major decision point in / remarks*
Commitment to division	*START* (yeast), restriction point (mammals)	Early in G_1	Activation	Protein kinase	Cultured animal cells
S phase initiation	—	End of G_1	Activation	Protein kinase	—
Initiation of cell division	Cell mass, DNA replication etc.	End of G_2	Prevention of M phase	M phase kinase	Naturally haploid organisms
	(*a*) Incomplete DNA replication*	End of G_2	Prevention of M phase	Phosphatase (Cdc25; activates Cdc2)	Activates M phase kinase
	(*b*) inadequate cell mass	G_2	Prevention of M phase	Protein kinase (Wee 1, Mik 1)**	Inactivates M phase kinase
	(*c*) DNA damage	G_2	Prevention of M phase	RAD9 and six other genes of *S. cerevisiae*	—
Exit from cell division	Initiation of cell division	During M phase	Inactivation; destruction of cyclin	Unknown	—

*Unreplicated DNA equivalent to about ~1% of the genome blocks the onset of M phase.
**Either protein kinase phosphorylates Cdc2 (Tyr-15), thereby preventing its activation.

6.3.3. Quiescence

Early in G_1 phase, some cells may enter G_0 phase. ***G_0 phase*** is similar to G_1 phase, except for the fact that cells in G_0 phase are nondividing and they can not enter *S* phase. This phase, therefore, is call ***quiescence***, and G_0 cells are known as ***quiescent.*** Certain types of G_0 cells can be stimulated to begin division; in such cases, the G_0 cells enter cell cycle early in G_1 phase. However, in haploid organisms

like yeast and most fungi, cells become quiescent in the G_2 phase (rather than in G_1 phase).

6.4. CELL DIVISION CYCLE GENES

Those genes whose products are required specifically for cell cycle to proceed normally are called ***cell division cycle* (cdc) *genes***. A mutation in a *cdc* gene blocks cell cycle at a specific stage. In yeast, about 80 different *cdc* genes have been identified. The products of *cdc* genes perform one of the following two functions: (1) they are needed for structural purposes, *e.g.*, enzymes of DNA replication, or (2) they are involved in the regulation of cell cycle. The *cdc* genes were identified from temperature sensitive *cell division cycle mutants*. In these mutants, normal cell division occurs at 23°C, but at 36°C cell cycle is blocked at specific stages. Yeast *cdc* mutants have been used to identify homologous *cdc* genes of higher eukaryotes by employing gene cloning strategies. Further, analysis of these mutants has also allowed the identification of checkpoints.

6.5. FACTOR REGULATING M PHASE INITIATION

The entry of cells into *S* phase and *M* phase is regulated by specific activators, which are protein kinases. Information on the nature of these activators was obtained from the following studies: (1) cell division in heterokaryons of amimal cells, and (2) maturation of *Xenopus* oocytes.

6.5.1. Animal Cell Heterokaryons

A ***heterokaryon*** is a cell that contains that two (or more) nuclei, usually of different genotypes, in a common cytoplasm. Heterokaryons are produced by fusing two or more cells with the help of, usually, PEG (polyethylene glycol). When a cell in *S* phase is fused with a cell in G_1 phase, the G_1 phase nucleus is induced to enter *S* phase. But when an *S* phase cell is fused with a G_2 phase cell, the G_2 phase nucleus does not enter *S* phase. In contrast, when a cell in G_1, *S* or G_2 phase is fused with a cell in *M* phase, the G_1, *S* or G_2 nucleus is also induced to enter *M* phase (Table 6·2).

These findings allow the following conclusions. (1) An activator present in *S* phase cells induces G_1 cells to enter *S* phase. However, this activator has no effect on G_2 phase nuclei. (2) Similarly, an activator present in *M* phase cells induces G_1, *S*, or G_2 nuclei to enter *M* phase.

TABLE 6.2

Consequences of fusion of animal cells in different stages of cell cycle

Stage of cell cycle at the time of fusion			*Heterokaryon : Stage of cell cycle in nucleus of*	
Cell A		*Cell B*	*Cell A*	*Cell B*
S phase	+	*G1* phase	*S* phase	*S* phase
S phase	+	*G2* phase	*S* phase continues; *G2* phase shortened	*G2* phase
G1/S/G2 phase	+	M phase	*M* phase	*M* phase
G1 phase	+	*G2* phase	*G1* phase	*G2* phase

6.5.2. Oocyte Maturation Promoting Factor

Immature oocytes of *Xenopus* are arrested at the start of meiosis that is equivalent to G_2 phase of mitosis. But when eggs are laid, they are arrested in the second division of meiosis at a stage that is equivalent to the *M* phase of somatic

cells. When cytoplasm from eggs is injected into immature oocytes, the latter are induced to enter meiotic division and undergo maturation. Therefore, the factor present in egg cytoplasm that induces the maturation of oocytes was called ***maturation promoting factor*** or ***MPF.***

When *Xenopus* eggs are fertilized, they undergo 11 mitotic divisions in quick succession. The first division takes ~ 90 minutes, while the subsequent divisions takes only ~30 min each. Thus each division consists of only *S* and M phases, and G_1 and G_2 phases are completely absent. The level of MPF activity peaks at the start of each division, and it disappears as the *S* phase begins.

M Phase Kinase (1)
- An activator that regulates entry of cells into *S* and *M phases.*
- *Xenopus* eggs have a oocyte maturation promoting factor (MPF).
- MPF activity peaks at cell division but drops to zero before *S* phase in *Xenopuos* young embryos.
- MPF induces somatic cells to enter M phase.
- MPF is now called M phase kinase.
- MPF has been purified; it has one p^{34} and one p^{45} subunits.

6.6. M PHASE KINASE

MPF also induces somatic cells to enter M phase; therefore, it is now called ***M phase promoting factor*** (in the place of *maturation promoting factor*). MPF has been isolated from *Xenopus* eggs; it is a dimer protein with protein kinase activity. Therefore, MPF is now known as ***M phase kinase***, which is the most commonly used name. MPF has two subunits designated as p^{34} (catalytic subunit) and p^{45} (regulatory subunit); the superscripts 34 and 45 indicate the molecular weights in kD (kilo Dalton) of the repective subunits.

6.6.1. Catalytic Subunit

The ***catalytic subunit*** is so called because it carries out phosphorylation of the target proteins. (A ***protein kinase*** is an enzyme that catalyzes phosphorylation of proteins.) In *Xenopus*, frogs and mammals, this subunit is designated as p^{34}; this subunit is known as CDC28 *in Saccharomyces cerevisiae*, and as Cdc2 in

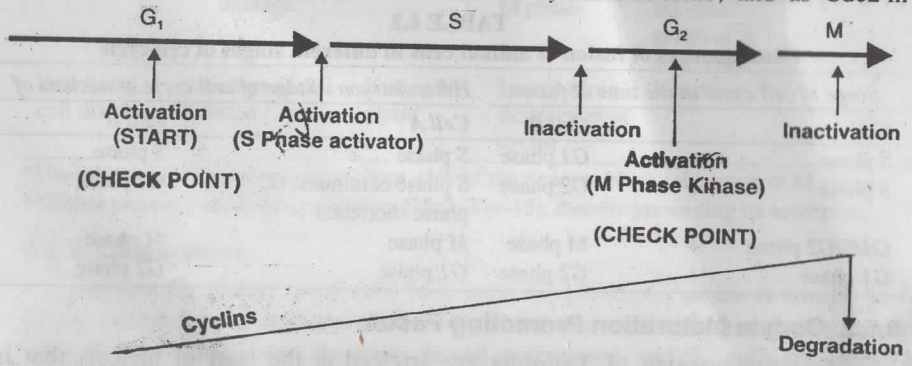


Fig. 6.2. A diagrammatic summary of the various activation and inactivation events that regulate somatic cell cycle, and the two important checkpoints in the cycle.

Schizosaccharomyces pombe. The level of p^{34} remains constant throughout the cell cycle and is in excess of p^{45}. During interphase, the catalysic subunit is present in an inactive form; it is converted into an active form toward the end of G_2 phase (Fig. 6·2).

The protein kinases that regulate the initiation of *M* phase and *S* phase have the same catalytic subunit in the yeasts *S. cerevisiae* and *S. pombe* (Table 6.3). But in mammals and frogs, *M* phase kinase has p^{34} as its catalytic subunit, while CDK2 and CDK4 constitute the catalytic subunits of the protein kinases that regulate the onset of *S* phase.

TABLE 6·3
The kinases that regulate entry of cells into M phase (M phase kinase) and S phase

Organism	*M phase kinase*	*Protein kinase regulating the onset of S phase*
	Catalytic Submit	
Mammals, frogs	P^{34}	CDK2 and CDK4
S. cerevisiae	CDC28	CDC28
S. pombe	Cdc2	Cdc2
	Regulatory Submit	
Mammals, frogs	Cyclins *A*, B_1, B_2	Cyclins *A*, *C*, D_1, *D*2, *D*3, *E*
S. cerevisiae	CLB 1 – CLB 4 (B – Like)	CLN 1 – CLN 3
S. pombe	Cdc13 (B-like)	Cig2

6.6.2. Regulatory subunit

The subunit p^{45} functions as the regulatory subunit of M phase kinase in mammals and frogs. This subunit is essential for the catalytic action of p^{34}, and allows M phase kinase to act on the specific target proteins. In animals, this subunit accumulates continuously through G_1, *S* and G_2 phases of cell cycle, and is destroyed abruptly by proteolysis following the onset of mitosis (Fig. 6.2). In view of this cyclical change in its concentration, this subunit was named as ***cyclin.*** Cyclin destruction during mitosis eliminates M phase kinase activity; this is essential for the cells to exist M phase. (If M phase kinase activity were maintained at the level it is present at the start of mitotic phase, the cells will be permanently locked in the M phase.)

M Phase Kinase (2)
- It is a protein kinase, *i.e.,* it phosphorylates target proteins.
- p^{34} is catalytic subunit (in animals; in yeast it is Cdc2 in *S. pombe* and CDC28 in *S. cerevisiae*).
- p^{34} level remains constant throughout the cell cycle.
- p^{45} is regulatory subunit in animals; in *S. pombe,* it is Cdc13, and is *S. cerevisiae* it is CLB1- CLB4; it is essential for p^{34} function.
- p^{45} level increases during *G1*, *S* and *G2.*
- p^{45} is degraded abruptly by proteolysis during early M phase.
- Regulatory subunit is commonly call *cyclin.*

In case of yeast, however, cyclins are not subject to abrupt proteolysis. Yeast cyclins have a rather short half life, *e.g.*, ~15 min for CLN2 of *S. cerevisiae*. As a result, yeast cyclins are unable to increase in concentration by gradual accumulation. But they do show marked increase in concentration just prior to the onset of *START* and of M phase. These increases are the result of increased rates of transcription of

Cyclins

- They form the regulatory subunit of M phase kinase and of kinase regulating the onset of *S* phase.
- They are of two types :
  1. *G_2 cyclins* : component of *M phase kinase;* cyclins *A*, B1 and B2 in animals, Cdc13 in *S. pombe*, CLB1-4 in *S. cerevisiae.*
  2. *G1 cyclins* : component of the kinase regulating onset of *S* phase; cyclins A, C, D1 - 3 and E in animals, Cig2 in *S. pombe*, CLN1- 3 in *S. cereivsiae.*
- Mechanism of regulation of cyclin concentration:
  1. *Gradual accumulation—abrupt proteolytic degradation* : cyclins A, B1 and B2.
  2. *Enhanced transcription rate* : cyclins *C, D*1 -3, and *E*, Cdc 13, Cig2, CLB1-4, and CLN1-3

cyclin genes. The same mechanism seems to be responsible for the increased levels of cyclins *C, D*1, *D*2, *D*3 and *E* of mammals and frogs; this occurs just before the or set of *S* phase.

Different cyclins are involved in kinase action regulating the onset of *S* and M phases; these cyclins are often called ***G_1 cyclins*** and ***G_2 cyclins***, respectively. In case of animals, cyclins A, B1 and B2 serve as *G_2 cyclins*, and are associated with M phase kinase. The concentrations of these cyclins are regulated by gradual accumulation and abrupt hydrolysis. These cyclins contain a sequence of ~150 amino acids, which is often called ***cyclin box***. They also contain a short sequence near their N-terminus, which is necessary for their proteolysis; this sequence is known as ***cyclin destruction box.*** In addition, polypeptide ubiquitin becomes associated with these cyclins to mark them as targets of proteolysis.

In *S. pombe,* M phase kinase is associated with a single type of regulatory subunit; this subunit is similar to cyclin B of animals and is encoded by gene *cdc*13. But in *S. cervisiae* four different types of regulatory subunits, *viz.*, CLB1, CLB2, CLB3 and CLB4, are associated with M phase kinase. These subunits are cyclin B-like, and any one of them is enough to ensure normal M phase kinase function.

The function of G_1 cyclins is performed by subunits CLN1, CLN2 and CLN3 in *S. cerevisiae*; these cyclins associate with the protein kinase that regulates the entry of cells into *S* phase. In *S. pombe*, polypeptide Cig2 functions as G_1 cyclin, while in animals cyclins A, C, D1, D2, D3 and E serve this function. The periodic accumulation of CLN1 – CLN3, Cig2, and cyclins C, D1 – D3 and E seems to be regulated at the level of transcription.

6.7. ACTIVATION / INACTIVATION OF M PHASE KINASE

In case of mammals and frogs, activation and inactivation of M phase kinase involves the followings: (1) phosphorylation of certain amino acids, (2) dephosphorylation of certain other amino acids and (3) proteolytic destruction of cyclins.

M Phase Kinase: Activation

- Accumulates as p^{34}-p^{45} dimer in an inactive form.
- Inactive form has P at Thr-14 and Tyr-15 of p^{34}.
- Activation occurs towards the end of *G2* phase by two events :
  1. *Dephosphorylation*: removal of P from Thr-14 and Tyr-15.
  2. *Phosphorylation*: Thr-161 is phosphorylated.
  (Both events are essential for activation)
- Activated kinase induces M phase.

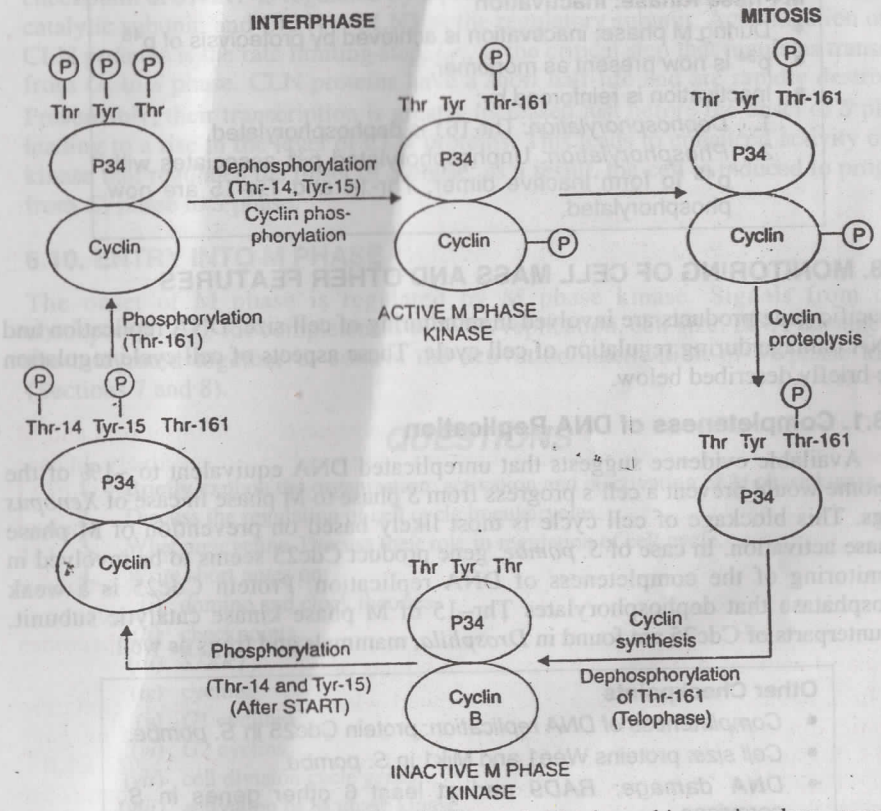


Fig. 6.3. Cyclic activation and inactivation of M phase kinase involves specific phosphorylation and dephosphorylation events and cyclin destruction. Cyclins A and B are degraded during metaphase and anaphase, respectively. Proteins are degraded by *anaphase promoting complex* (APC) that has 8 subunits and functions as an E3 ubiquitin ligase.

6.7.1. Inactivation of M Phase Kinase

In mammalian cells, M phase kinase is inactivated after the onset of M phase: this is achieved by proteolytic degradation of the cyclin component of the active

dimer. This leaves the catalytic unit, p^{34}, as a monomer. Presumably, the Thr-161 of p^{34} is now dephosphorylated. As cyclins accumulate during G_1, S and G_2, they associate with the non-phosphorylated p^{34} subunits to form inactive dimer molecules. Following dimer formation, Thr-14 and Tyr-15 of p^{34} are phosphorylated to reinforce inactivation of M phase kinase. The M phase kinase remains in this inactive form till its activation at the end of G_2. The M phase inactivation involves the following events (Fig. 6·3).

1. Destruction of the cyclin component of M phase kinase dimer.
2. Dephosphorylation of Thr–161 of p^{34} subunit (now present as monomer).
3. Phosphorylation of Thr–14 and Tyr–15 of p^{34} subunit. (This occurs after p^{34} associates with cyclin to form inactive dimers.)

M Phase Kinase: Inactivation

- During M phase; inactivation is achieved by proteolysis of p^{45}.
- p^{34} is now present as monomer.
- Inactivation is reinforced by:
  1. *Dephosphorylation:* Thr-161 is dephosphorylated.
  2. *Phosphorylation*: Unphosphorylated p^{34} associates with p^{45} to form inactive dimer; Thr-14 and Tyr-15 are now phosphorylated.

6.8. MONITORING OF CELL MASS AND OTHER FEATURES

Specific gene products are involved in monitoring of cell size, DNA replication and DNA damage during regulation of cell cycle. These aspects of cell cycle regulation are briefly described below.

6.8.1. Completeness of DNA Replication

Available evidence suggests that unreplicated DNA equivalent to ~1% of the genome would prevent a cell's progress from *S* phase to M phase in case of *Xenopus* eggs. This blockage of cell cycle is most likely based on prevention of M phase kinase activation. In case of *S. pombe*, gene product Cdc25 seems to be involved in monitoring of the completeness of DNA replication. Protein Cdc25 is a weak phosphatase that dephosphorylates Thr–15 of M phase kinase catalytic subunit. Counterparts of Cdc25 are found in *Drosphila*, mammals and frogs as well.

Other Checkpoints

- *Completeness of DNA replication:* protein Cdc25 in *S. pombe.*
- *Cell size*: proteins Wee1 and Mik1 in *S. pombe.*
- *DNA damage: RAD9* and at least 6 other genes in *S. cerevisae.*
- Checkpoints for DNA replication and cell size regulate M phase kinase activation/inactivation.

6.8.2. Regulation of Cell size

The gene products involved in regulation of cell size are Wee1 and Mik1 in case of *S. pombe*; these proteins also act by regulating the activation/inactivation of M phase kinase. Cells mutant for *wee*1 and *mik*1 genes divide at a much smaller size

than do normal cells. Protein Wee1 is a protein kinase that inactivates the catalytic subunit by phosphorylating Thr–15; protein Mik1 performs the same function as Wee1.

6.8.3. Monitoring DNA Damage

S. cerevisiae cells are blocked in G_2 if they have damaged DNA. Inactivation of gene *RAD9* by mutation allows cells to enter *M* phase even when they contain DNA damage. Thus *RAD9* gene product acts as a checkpoint for monitoring DNA damage. RAD9 protein functions together with products of 6 other genes.

6.9. INITIATION OF *S* PHASE

The regulation of onset of *S* phase is best understood in case of *S. cerevisiae*. The checkpoint at *START* is regulated by a protein kinase. This kinase has CDC28 as the catalytic subunit, and CLN1 – CLN3 as the regulatory subunit. Accumulation of the CLN proteins is the rate limiting step, *i.e.*, is the critical step that regulates transition from G_1 to *S* phase. CLN proteins have a short half-life and are rapidly destroyed. Presumably, their transcription is greatly increased just before the onset of *S* phase, leading to a rise in the level of CLN proteins. This leads to enhanced activity of the kinase that regulates the onset of *S* phase; as a result, the cell is induced to progress from G_1 phase to *S* phase.

6.10. ENTRY INTO M PHASE

The onset of M phase is regulated by M phase kinase. Signals from other checkpoints, *e.g.*, for completeness of DNA replication, cell size, DNA damage etc., are integrated together to control the activation/inactivation of M phase kinase (Sections 7 and 8).

QUESTIONS

1. Briefly explain the organization, activation and inactivation of *M* phase kinase.
2. Discuss the regulation of cell cycle in eukaryotes.
3. What are cyclins. Discuss their role in regulation of cell cycle.
4. Write short notes on:
   (*i*) domino and clock theories,
   (*ii*) checkpoints,
   (*iii*) *START*,
   (*iv*) cyclins,
   (*v*) *G*1 cyclines,
   (*vi*) *G*2 cyclins,
   (*vii*) cell division cycle genes,
   (*viii*) activation of M phase kinase,
   (*ix*) inactivation of M phase kinase,
   (*x*) checkpoint for DNA replication, cell size and DNA damage.

7

Chromosomal Aberrations
I. Structural Aberrations

7.1. INTRODUCTION

The somatic (2n) *and gametic* (n) *chromosome numbers of a species ordinarily remain constant.* This is due to the extremely precise mitotic and meiotic cell divisions. Somatic cells of a diploid species contain two copies of each chromosome, which are called ***homologous chromosomes***. Their gametes, therefore, contain only one copy of each chromosome, that is, they contain one chromosome complement or genome. *Chromosomes of a* ***genome*** *are distinct from one another in their morphology and/or gene content.* ***Members of a single genome do not show a tendency to pair with each other.*** *Each chromosome of a genome contains a definite number and kind of genes, which are arranged in a definite sequence. Homologous chromosomes contain an identical number and kind of genes in the same sequence.*

Chromosomal Aberrations

- Deviation in chromosome structure on number.
- *Types*: (*i*) structural
  (*ii*) numerical (heteroploidy).
- *Structural aberrations*: altered number or sequence of genes present in chromosomes. Types :
  (*i*) deletion or deficiency,
  (*ii*) duplication or repeat,
  (*iii*) inversion,
  (*iv*) translocation.

Occasionally, spontaneous (without any known causal factor) variations in chromosome number or structure do arise in nature. These variations are called ***chromosomal aberrations***. Chromosomal aberrations may be grouped into two broad classes: (1) structural, and (2) numerical (described in Chapter 8).

7.2. STRUCTURAL CHROMOSOMAL ABERRATIONS

In this class are included those chromosomal aberrations, which alter the chromosome structure, i.e., *the number, the sequence or the kind of genes present in chromosome*(*s*). There are four common types of structural aberrations: (1) deletion or deficiency, (2) duplication or repeat, (3) inversion, and (4) translocation (Table 7·1). *Deletions (a decrease) and duplications (an increase) alter the number of genes present, inversions change the gene sequence, while translocations affect the kind of genes located in the affected chromosomes.*

7.3. ORIGIN OF STRUCTURAL ABERRATIONS

All structural aberrations are produced following chromosome breakage. Chromosome breakage occurs spontaneously in a low frequency (Ca.1%) of the cells in almost all the tissues studied. The cause of spontaneous chromosome breakage is not definitely known.

TABLE 7·1
A Summary of Different Structural Changes in Chromosomes

Term	*Type of aberration*	*Remarks*
Deletion	Loss of a segment of a chromosome.	Produces pseudodominance for recessive alleles located in the lost segment.
1. *Terminal deletion*	Lost segment includes telomere.	Believed to be rare.
2. *Interstitial deletion*	A segment between telomere and centromere is lost.	Most deletions are of this type.
Duplication (*Repeat*)	A chromosome segment present in more than two copies in the same nucleus.	According to Ohno (1970), duplications are the source of all new genes, and thus the basis of organic evolution.
1. *Tandem duplication*	The additional chromosome segment is located just after the normal segment, gene sequence being the same.	
2. *Reverse duplication*	Same as above but the gene sequence of the additional segment is inverted.	
3. *Displaced duplication*	The additional segment located in the same chromosome but away from the normal segment.	
4. *Translocation duplication*	The additional segment located in a nonhomologous chromosome.	
Inversion	A chromosome segment contains genes in a sequence, which is the reverse of the normal.	Produces characteristic dicentric bridge and acentiric fragment at AI in inversion heterozygotes.
1. *Paracentric inversion*	The inverted segment does not contain centromere.	
2. *Pericentric inversion*	The inverted segment contains centromere.	May change centromere location.
Translocation	Chromosome segment(s) integrated into nonhomologous chromosome(s).	The main mechanism for change in chromosome number and morphology in nature.
1. *Simple translocation* (shift)	A segment of a chromosome integrated into a nonhomologous chromosome.	May lead to translocation duplication.
2. *Reciprocal translocation* (exchange)	A segment of chromosome integrated into a nonhomologous chromosome from which a segment is integrated into the first one.	Characteristic ring or chain of four or more chromosomes at MI in the translocation heterozygotes.

Chromosome breakage is induced in a relatively high frequency by several radiations, *e.g.*, X-rays, gamma-rays, α-rays, β-rays, neutrons, chemical agents, *e.g.*,

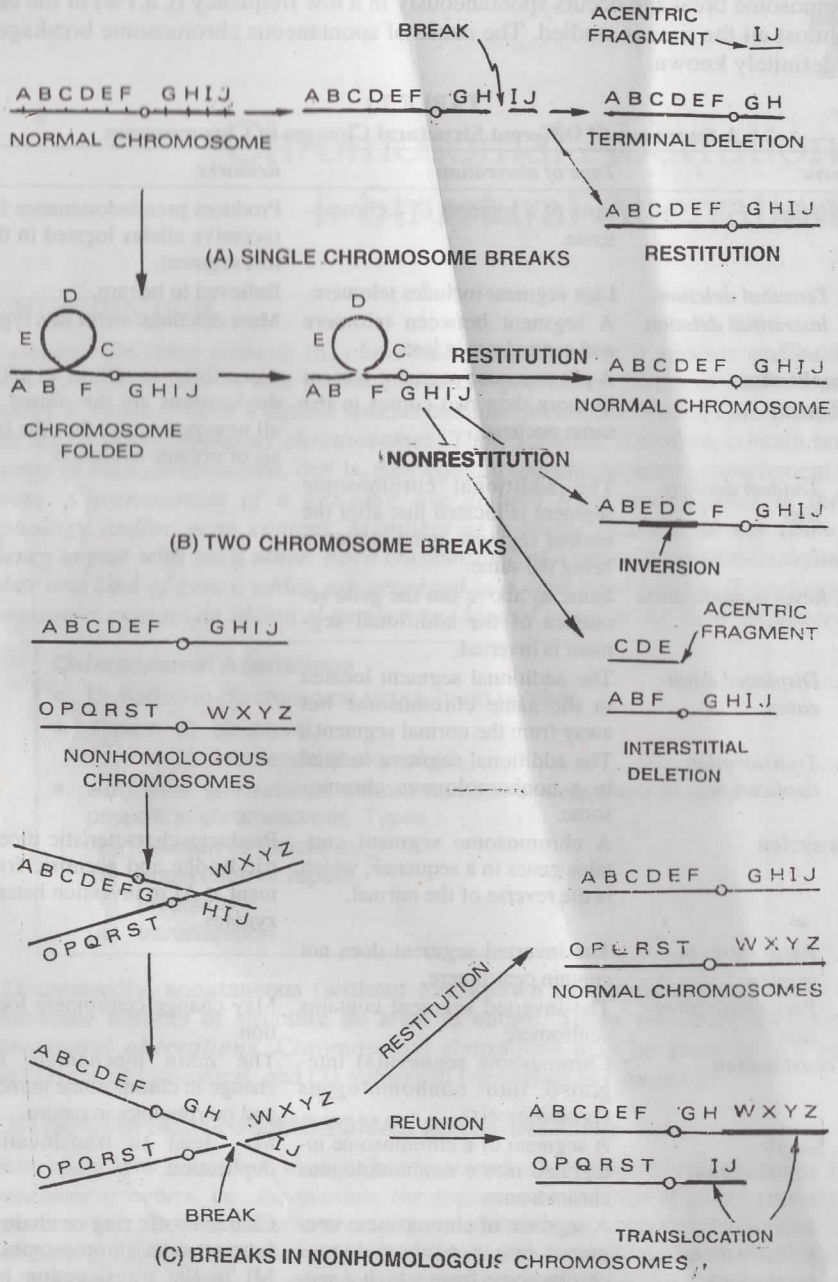


Fig. 7·1. A schematic representation of the possible modes of origin of structural aberrations in chromosomes.

ethylmethane sulphonate, base analogues, many insecticides, herbicides and fungicides etc., and by several viruses, *e.g.*, measles virus. In addition, some genes are also known to induce chromosome breakage, *e.g.*, *Ac-Ds* of maize described by McClintock and other transposable elements of eukaryotes, and some other gene mutations.

7.3.1. Terminal Deficiency

Since broken chromosome ends lack telomere, they are highly unstable, and are prone to unite with other damaged ends. When a break occurs in a chromosome, the two broken ends thus produced often join with each other producing the same original chromosome; this is known as ***restitution*** (Fig. 7.1A). Sometimes, the two broken ends may heal and the acentric fragment thus produced is generally lost, producing a ***terminal deficiency.*** This is always associated with the production of an acentric fragment. An ***acentric fragment*** does not have centromere. Such fragments are ultimately lost as they are digested by exonucleases.

7.3.2. Interstitial Deficiency and Inversion

Most structural aberrations, however, involve two breaks. A chromosome may be folded onto itself, and the two breaks may occur at or close to the point where the chromosome passes over itself. Thus the following chromosome segments will be produced: (1) segment *AB* containing the telomere, (2) segment *CDE* (acentric segment), and (3) centric fragment (containing the centromere) FGHIJ (Fig. 7.1B). Often the broken ends of the three segments will join in their original sequence producing the normal chromosome ABCDEFGHIJ (***restitution***). But there may be ***nonrestitution*** or ***reunion***, that is, the broken ends of the chromosome may unite in a pattern other than their original sequence. This would produce structural aberrations. For example, the segment *CDE* may be lost as an acentric fragment, and the end *B* of segment *AB* may unite with end *F* of the centric fragment. This will produce ***interstitial deficiency*** for *CDE*. Alternatively, the sequence of *CDE* may be reversed, *i.e.*, inverted. Now the end *E* of *CDE* reunites with end *B* of *AB*, and end *C* of *CDE* reunites with end *F* of the centric segment; this would produce ***inversion*** for *CDE* (Fig. 7.1B).

7.3.3. Translocation

When two nonhomologous chromosomes pass over each other, breaks may occur in them at or close to the point of their contact. A reunion may occur

Deletion

- Discovered by Bridges in 1917 from genetic studies.
  1. *Terminal*: terminal past broken and lost; one break; rare.
  2. *Interstitial*: two breaks; internal chromosome segment missing (Syn., intercalary).
- Recessive lethals (mostly).
- Many cause dominant genotypic effects, *e.g.., Notch* wings.
- *Pseudodominance* in the deleted segment.
- Lack of recombination in the deleted segment.
- Loop formed in pachytene in heterozygotes.

between the centric segment of one chromosome with the acentric fragment of the other chromosome and *vice-versa*. Such a reunion will generate a ***reciprocal translocation*** between the two chromosomes (Fig. 7.1C).

7.4. DELETION

Loss of a chromosome segment is known as ***deletion*** or ***deficiency***. When the missing segment contains a telomere of the affected chromosome, it is called ***terminal deletion*** (Fig. 7.2 A,B). *Such deletions are believed to be uncommon.* In most deficiencies, however, the missing segment does not contain a telomere; such deletions are known as ***interstitial*** or ***intercalary*** (Fig. 7.2 A,C).

Deletion was the first structural aberration detected by Bridges in 1917 from his genetic studies on the X chromosome of *Drosophila*. In 1938, Muller postulated that the loss of telomere makes a chromosome end unstable. Therefore, all deficiencies that are found in nature must be intercalary.

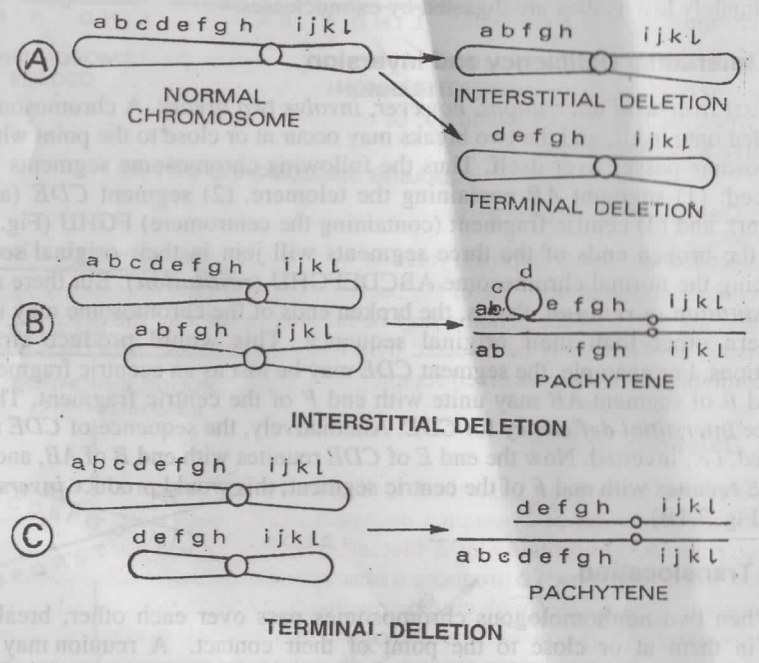


Fig. 7.2. (A) Terminal and interstitial (interclalary) deletions. Pachytene chromosome configurations in (B) terminal and (A) interstitial deletions.

7.4.1. Genetic Effects

Deletions generally produce striking genetic and morphological/physiological effects. (1) *The loss of a chromosome segment has a harmful effect on diploid organisms.* (2) *Most deficiencies act as recessive lethals.* Larger deficiencies may produce dominant lethal effect. (3) Individuals heterozygous for a deficiency contain only one copy of the deleted segment. Thus, *the deleted segment will be in hemizygous state. Therefore, recessive alleles of the genes located in the deleted segment will express themselves in deficiency heterozygotes as if these alleles were dominant. This phenomenon is known as* **pseudodominance**. (4) *Crossing over is completely absent in the deleted region* of a chromosome since this region is present in only one copy in the deletion heterozygotoes. Finally, (5) *many deletions produce a distinct observable morphological effect.* In *Drosophila,* several deficiencies produce dominant phenotypic as well as recessive lethal effects, *e.g., Beaded, Delta, Gull, Minute, Notch, Notopleural, Plexate* etc. These deficiencies were recognised as distinct genes and are mapped as such on the linkages map of *Drosophila.*

7.4.2. Cytological Detection

Cytological detection of deficiencies for relatively larger chromosome segments presents little difficulty. *During pachytene, the unpaired segment of the normal chromosome of an intercalary deletion heterozygote produces a characteristic loop* (Fig. 7.2B). *In case of terminal deletion heterozygotes, the segment toward one end of the normal chromosome remains unpaired.* Both these configurations have been described in maize by Creighton and McClintock, and have been observed in the giant chromosomes of *Drosophila.*

Detection of smaller deletions, however, is very difficult in most organisms except where giant chromosomes are available for cytological examination. In most eukaryotes, such deletions would not be recognised and they would be classified as recessive lethal mutations.

7.4.3. Deletions in Prokaryotes

Deletions are found in prokaryotes as well, *e.g., E. coli,* phage T_4, phage λ, etc. In *E. coli*, deletions of upto 1% of the bacterial chromosome are known. In phage λ, however, upto 20% of the genome may be missing in some of the deletions. In these organisms, deletions may be detected by a loss of a function associated with a lack of recombination in the concerned region.

7.4.4. Deletions in Man

Deletions have been described in human beings. Each deletion appears to produce a distinct set of symptoms or 'syndromes'. The group of symptoms that characterizes an abnormality or a disease is known as a ***syndrome***. A specific deletion in chromosome 22 produces a chromosome called '*Philadelphia* 22'. This chromosome is associated with chronic myelogenous leukaemia. Another deficiency in chromosome 5 produces '*cri-du-chat*' (cry-of-cat) syndrome. Individuals suffering from this syndrome produce a characteristic mewing catlike cry during childhood. They also have unique facial features and exhibit severe physical and mental retardation.

Duplication

- Additional segment present; Types: (*i*) tandem; (*ii*) reverse tandem; (*iii*) displaced; (*iv*) translocation.
- Produced by unequal crossing over / breakage and reunion.
- Loop at pachytene in heterozyotes.
- Distinct phenotypic effects, *e.g., Bar* eye.
- Source of new genes (in combination with mutation).

7.5. DUPLICATION

The presence of an additional chromosome segment, as compared to that normally present in a nucleus, is known as ***duplication***. In a diploid organism, presence of a chromosome segment in more than two copies per nucleus is called ***duplication***.

The extra chromosome segment may be located immediately after the normal segment in precisely the same orientation, producing ***tandem duplication*** (Fig. 7.3). When the gene sequence in the extra segment of a tandem duplication is in reverse order, *i.e.*, is inverted (Fig. 7.3), it is known as ***reverse tandem duplication***. In some cases, the extra segment may be located in the same chromosome, but away from the normal segment; they are termed as ***displaced duplications***. In ***translocation duplications,*** the additional chromosome segment is located in a nonhomologous chromosome.

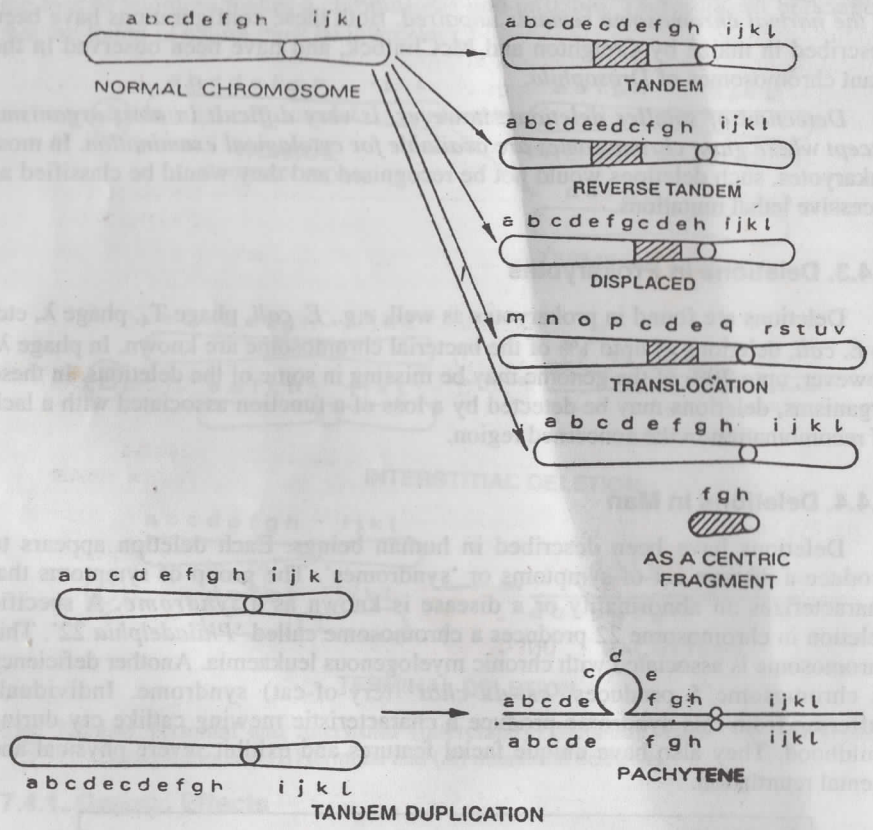


Fig. 7.3. Different types of duplications, and pachytene chromosome configuration in a tandem duplication heterozygote (an individual having one normal and one duplicated chromosome).

7.5.1. Origin

Origin of duplications involves chromosome breakage and reunion of a chromosome segment with its homologous chromosome. As a result, one of the two homologues involved in the production of a duplication ends up with a deficiency,

while the other has a duplication for the concerned segment. Another phenomenon, known as ***unequal crossing over,*** also leads to exactly the same consequences for small chromosome segments. For example, duplication of the band *16A* of *X* chromosome of *Drosophila* produces *Bar* eye. This duplication is believed to originate due to unequal crossing over between the two normal *X* chromosomes of females (Fig. 7.4). One of the two *X* chromosomes involved in unequal crossing over receives two *16A* bands (*bar* chromosome), while the other receives none. Similarly, unequal crossing over may occur between two *X* chromosomes that are duplicated for the band *16A*. This would produce one normal *X* (having one *16A* band) and one *ultrabar X* chromosome (having three *16A* bands).

7.5.2. Cytology

Relatively larger duplications are detectable in heterozygous state due to the presence of loops during pachytene stage (*tandem duplications*). In the giant chromosomes of Diptera, duplications are gnerally observable as loops in duplication heterozygotes (*tandem* or *displaced duplications*) or as cross-pairings between segments of non-homologous chromosomes (*translocation* duplications). *Relatively smaller duplications are detectable as repeat occurrences of bands in salivary gland chromosomes,* which have a highly characteristic banding pattern.

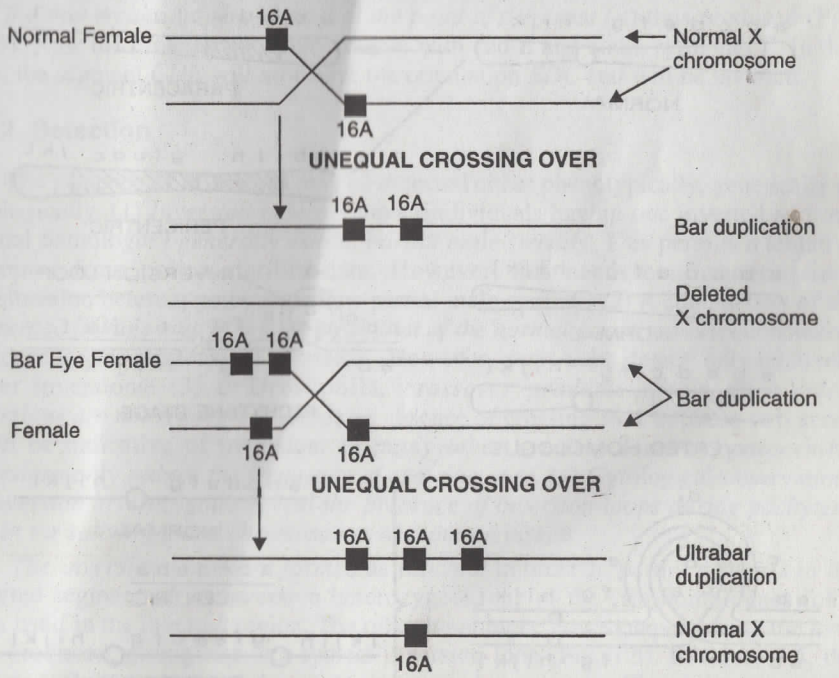


Fig. 7.4. Production of *bar* eye duplication (*16A* band of the *X* chromosome) of *Drosophila.* Unequal crossing over between two *bar X* chromosomes produces one *ultrabar* and one normal *X* chromosome.

7.5.3. Genetic Effects

In general, duplications do not produce as drastic consequences as do deletions. *But some duplications may have distinct phenotypic effects and may be regarded as*

mutant alleles, e.g., Bar *duplication in* Drosophila *(Fig. 7.4). It is postulated that duplicate, complementary etc. genes and multiple factors (polygenes) have originated due to duplications. Ohno (1970) has postulated that the increase in DNA content per cell accompanying evolution and the* ***origin of new genes*** *with distinct functions is solely due to duplications.* According to him, all new loci, which have developed during evolution owe their origin to gene duplication. Ohno has presented considerable evidence in support of this hypothesis.

7.6. INVERSION

When a segment of a chromosome is oriented in the reverse direction (Fig. 7.5), such a segment is said to be ***inverted***, and the phenomenon is termed as ***inversion***. Obviously, the gene sequence in an inverted segment is exactly the opposite of that in its normal homologous (noninverted) segment. Therefore, inversions can be readily detected by a comparison of linkage maps of the normal and inverted chromosomes. The existence of inversion was first detected by Sturtevant and Plunkett in 1926 in this manner.

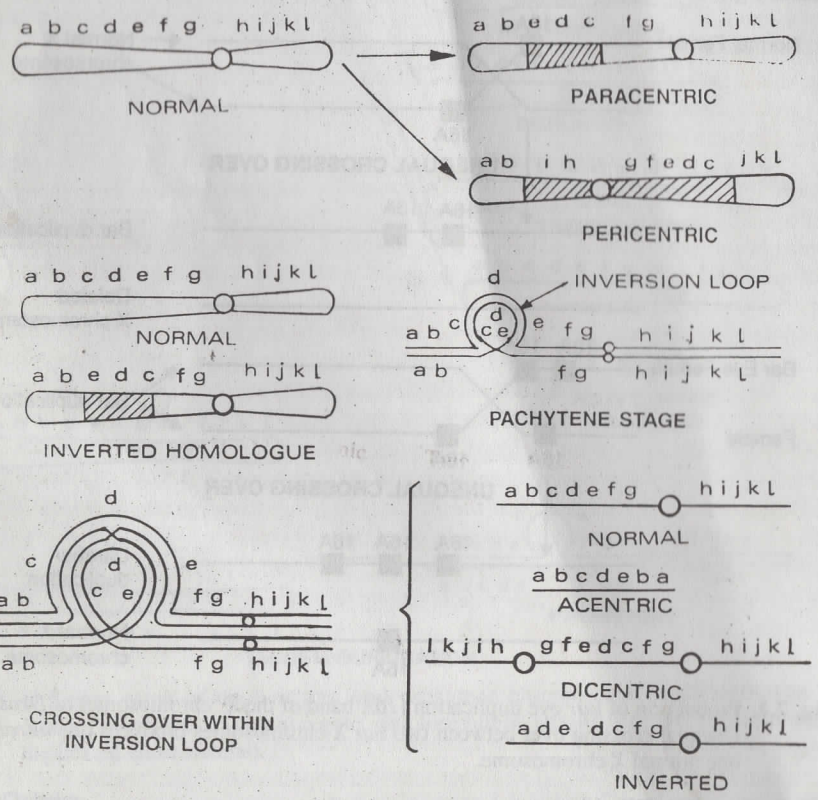


Fig. 7.5. Inverted chromosomes, and pairing and crossing over in an inversion heterozygote (paracentric inversion).

Inversion

- A chromosome segment present in reverse orientation.
- *Production*: two breaks; reunion with the internal segment in reverse orientation.
- Characteristic (*i*) inversion loop in hetrozyogote and (*ii*) partial pollen sterility.
- *Discovery*: Sturtevent and Plunkett in 1926 from inverted linage map.
- Quite frequent in natural populations.
- *Types*: (*i*) paracentric and (*ii*) pericentric.

7.6.1. Origin

By definition, the inverted segment is rotated by 180° as compared to its normal orientation. This would involve two chromosome breaks and reunion of the interstitial segment thus produced after it has rotated by 180° (Fig. 7.1B). This rotation may be achieved as follows. *A chromosome may become folded onto itself and the two breaks may be located at the point of the cross* (×) thus produced (Fig. 7.1B). Now the broken end *B* may reunite with end *E* and end *C* with end *F*. In this case, the segment *CDE* will now have the orientation *EDC* and will be inverted.

7.6.2. Detection

The presence of inversions may be detected either phenotypically, genetically or cytologically. (1) *Inversion heterozygotes* (individuals having one inverted and one normal homologue) *generally exhibit partial male sterility.* This permits a tentative inference from pollen sterility data. However, this needs confirmation since translocation heterozygotes also show partial male sterility. (2) *A comparison of the concerned linkage map of a strain with that of the normal strain* provides conclusive evidence for the existence of inversion. However, such data detect only relatively larger inversions. (3) *In* Drosophila, *crossover products within paracentric inversions are not recovered.* Hence an absence of crossing over between two genes could be indicative of inversion. *In many other species, however, paracentric inversions only reduce the frequency of crossing over. (4)* Cytological observations of *inversion heterozygotes reveal the presence of inversion loops during pachytene and in the salivary gland chromosomes of dipteran insects.*

The inversion loops are formed as follows. In order to permit synapsis in the inverted segment of an inversion heterozygote, one of the two homologues folds upon itself in the inverted region. The other homologue now synapses along the loop thus produced, giving rise to a typical inversion loop (Fig. 7.5). In figure 7.5, the inverted chromosome is shown to have folded onto itself, while its normal homologue simply follows the loop thus produced. The folded chromosome forming the loop could just as well be the normal chromosome; in such a case, the inverted homologue would follow the loop. *Inversion loops would be formed only when the inverted segments are large enough to form loops and permit chromosome pairing.*

(5) But inverted band patterns of salivary gland chromosomes make it possible to detect even small inversions, and that too in inversion homozygotes as well.

Therefore, inversions have been studied in a considerable detail in *Drosophila* species.

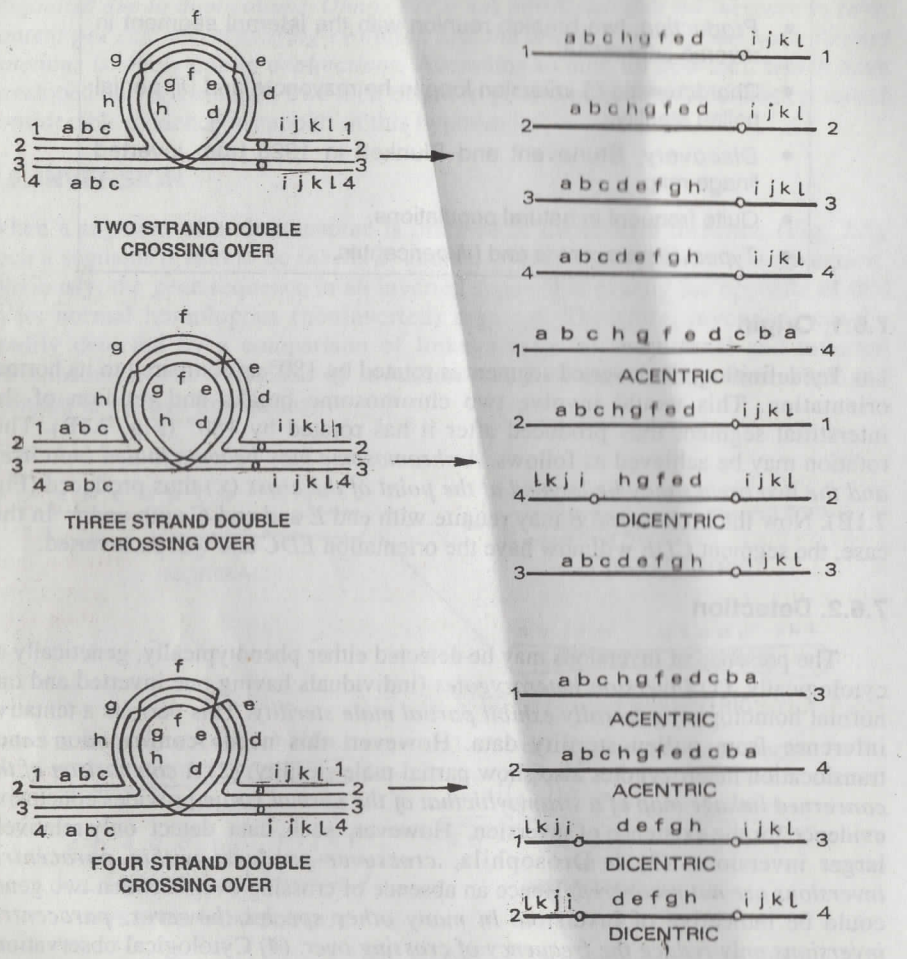


Fig. 7.6. Consequences of double crossing over within the inversion loop of a paracentric inversion. Note that only two-strand double crossing over produces normal cross-over chromosomes.

Paracentric Inversion

- Centromere not present in the inverted segment.
- A single crossing over within the loop produces at AI acentric fragment and dicentric bridge.
- Reduces recovery of crossover products (*Crossover suppressor, C*).
- Quite frequent in natural populations.

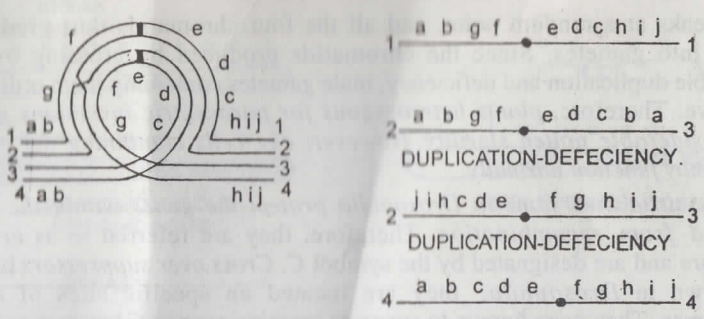


Fig. 7.7. Consequences of a single crossing over within the inversion loop of a pericentric inversion heterzygote. Note that the crossover products are duplicated for some region and deficient for some other region of the chromosome.

7.6.3. Types of Inversion

Inversions may be classified into two groups on the basis of the presence or absence of centromere within the inverted segment. In a ***paracentric inversion,*** the centromere is located outside the inverted segment (Fig. 7.5.). But a ***pericentric inversion*** contains the centromere within the inverted segment (Fig. 7.7). Inversion loops produced during synapsis in inversion heterozygotes are similar in appearance in both the cases (Figs. 7.5 and 7.7). However, the consequences of crossing over within the inversion loops are quite different in the two types of inversions. Paracentric inversions do not alter the morphology of the affected chromosomes, while pericentric inversions may alter the centromeric position (Fig. 7.8).

7.6.4. Crossing Over within Inversion Loop

A single crossing over within the inversion loop in the case of a paracentric inversion produces an acentric fragment and a dicentric chromatid from the two chromatids involved in the crossing over (Fig. 7.5). *As a result, a dicentric bridge along with an acentric fragment can be seen at AI in such cases; they are sure symptoms of paracentric inversion.* In most organisms, *e.g.,* maize, the dicentric

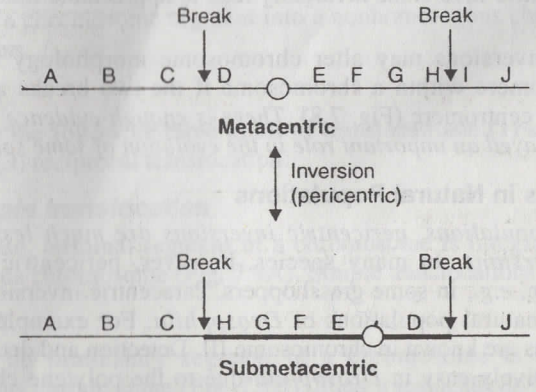


Fig. 7·8. The production of a submetacentric chromosome from a metacentric one and *vice-versa* through pericentric inversion.

bridge breaks at a random point, and all the four chromatids thus produced are included into gametes. Since the chromatids produced by crossing over have considerable duplication and deficiency, male gametes containing them ordinarily do not survive. Therefore, *plants heterozygous for paracentric inversions generally show considerable pollen sterility. However, egg cells containing the crossover products may function normally.*

Paracentric inversions in Drosophila protect the genes within the inverted segmented from recombination. Therefore, they are referred to as ***crossover suppressors*** and are designated by the symbol *C*. *Cross over suppressors* have been long known in *Drosophila;* they are located on specific sites of different chromosomes. They were known to suppress crossing over in *C* heterozygote (C^+ *C*) females. But in *C* homozygotes (*CC*) they had no effect. The fact that crossover suppressors are inversions, and the mechanism of their action was elucidated by Sturtevent and others during 1920's.

Double crossing over within a paracentric inversion loop leads to variable consequences depending upon the chromatids involved. A two-strand double crossing over produces two normal and two inverted chromatids (Fig. 7.6). But a three-strand double crossing over produces an acentric fragment and a dicentric chromatid similar to the single crossing over (Fig. 7.6). A four-strand double crossing over, on the other hand, yields two acentric and two dicentric chromatids.

Pericentric Inversion

- Centromere located within inverted segment.
- Crossing over produces monocentric products having duplication - deficiency.
- May change centromere position.
- Less frequent in natural populations; frequent in some species.

A single crossing over within the inversion loop of a ***pericentric inversion*** does not produce acentric and dicentric chromatids. *All the crossover products are monocentric but they have considerable duplications and deficiencies* (Fig. 7.7). As a result, male gametes carrying the crossover products are generally inviable. Therefore, pericentric inversions invariably lead to appreciable male sterility in the heterozygotes.

Pericentric inversions may alter chromosome morphology by shifting the position of centromere within a chromosome if the two breaks are not at equal distance from the centromere (Fig. 7.8). *There is enough evidence that pericentric inversions have played an important role in the evolution of some species.*

7.6.5. Inversions in Natural Populations

In natural populations, pericentric inversions are much less frequent than paracentric inversions. In many species, however, pericentric inversions are relatively common, *e.g.,* in some grasshoppers. Paracentric inversions appear to be very frequent in natural populations of *Drosophila*. For example, more than 31 different inversions are known in chromosome III. Detection and detailed analysis of inversions is relatively easy in *Drosophila* due to the polytene chromosomes. In other organisms, they may be equally frequent but their detection is generally not possible without a pachytene analysis of appropriate crosses.

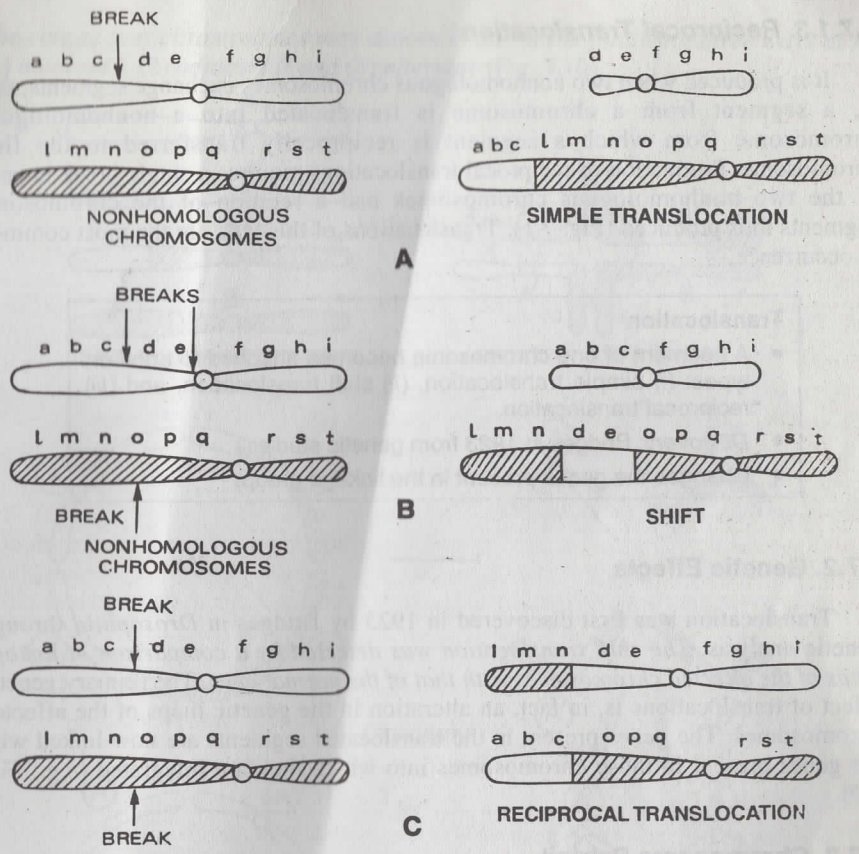


Fig. 7.9. Different types of translocation: (A) simple translocation, (B) shift, and (C) reciprocal translocation.

7.7. TRANSLOCATION

Integration of a chromosome segment into a nonhomologous chromosome is known as ***translocation.***

7.7.1. Types

There are the following three kinds of translocations: (1) simple translocation, (2) shift, and (3) reciprocal translocation.

7.7.1.1. Simple translocation

In this case, terminal segment of a chromosome is integrated at one end of a nonhomologous chromosome (Fig. 7.9A). Simple translocations are rather rare.

7.7.1.2. Shift

In shift, an intercalary segment of a chromosome is integrated within a nonhomologous chromosome (Fig. 7.9B). Such translocations (shifts) are known in natural populations of *Drosophila, Neurospora* etc.

7.7.1.3. Reciprocal Translocation

It is produced when two nonhomologous chromosomes exchange segments, that is, a segment from a chromosome is translocated into a nonhomologous chromosome from which a segment is reciprocally transferred to the first chromosome. Production of reciprocal translocations requires a single break in each of the two nonhomologous chromosomes and a reunion of the chromosome segments thus produced (Fig. 7.1). Translocations of this type are the most common in occurrence.

Translocation

- A segment of one chromosome becomes attached to another; types: (*i*) simple translocation, (*ii*) shift translocation, and (*iii*) reciprocal translocation.
- *Discovery*: Bridges in 1923 from genetic studies.
- Changes the genes present in the linkage group.

7.7.2. Genetic Effects

Translocation was first discovered in 1923 by Bridges in *Drosophila* through genetic analysis. *The shift translocation was detected by a comparison of linkage maps of the affected chromosomes with that of the normal ones.* The primary genetic effect of translocations is, in fact, an alteration in the genetic maps of the affected chromosomes. The genes present in the translocated segments are now linked with the genes present in those chromosomes into which they are integrated (*e.g.*, Fig. 7.9).

7.7.3. Chromosome Pairing

Cytology of reciprocal translocations has been investigated in a considerable detail. An individual having reciprocally translocated chromosomes may be either a translocation homozygote or a heterozygote. In a ***translocation homozygote,*** the two homologues of each of the two translocated chromosomes are identical in their gene content. As a result, they form normal bivalents, and there is no detectable cytogenetic peculiarity during meiosis. But there may be a change in chromosome morphology due to the translocation.

On the other hand, in a ***translocation heterozygote,*** one member from each of the two chromosome pairs is involved in a reciprocal translocation, while the remaining chromosome of the two concerned pairs is normal (Fig. 7.10). *Chromosome pairing in a translocation (reciprocal) heterozygote produces a peculiar + -shaped pachytene configuration. The point of the 'cross', i.e., the point where the four chromosomes converge, at pachytene represents the breakpoints in the chromosomes. Obviously, translocation heterozygotes of* Drosophila *will show cross-configurations in their salivary gland chromosomes.*

Crossing over occurs regularly as a result of which a ring of four (4) *chromosomes is seen at MI* (Fig. 7.10). *Whenever crossing over does not take place in one of the four pairing segments, a chain of four chromosomes is observed at MI.*

The ring of four chromosomes may disjoin in one of the following three ways at AI: (1) *adjacent I,* (2) *adjacent II and* (3) *alternate* (Fig. 7.10).

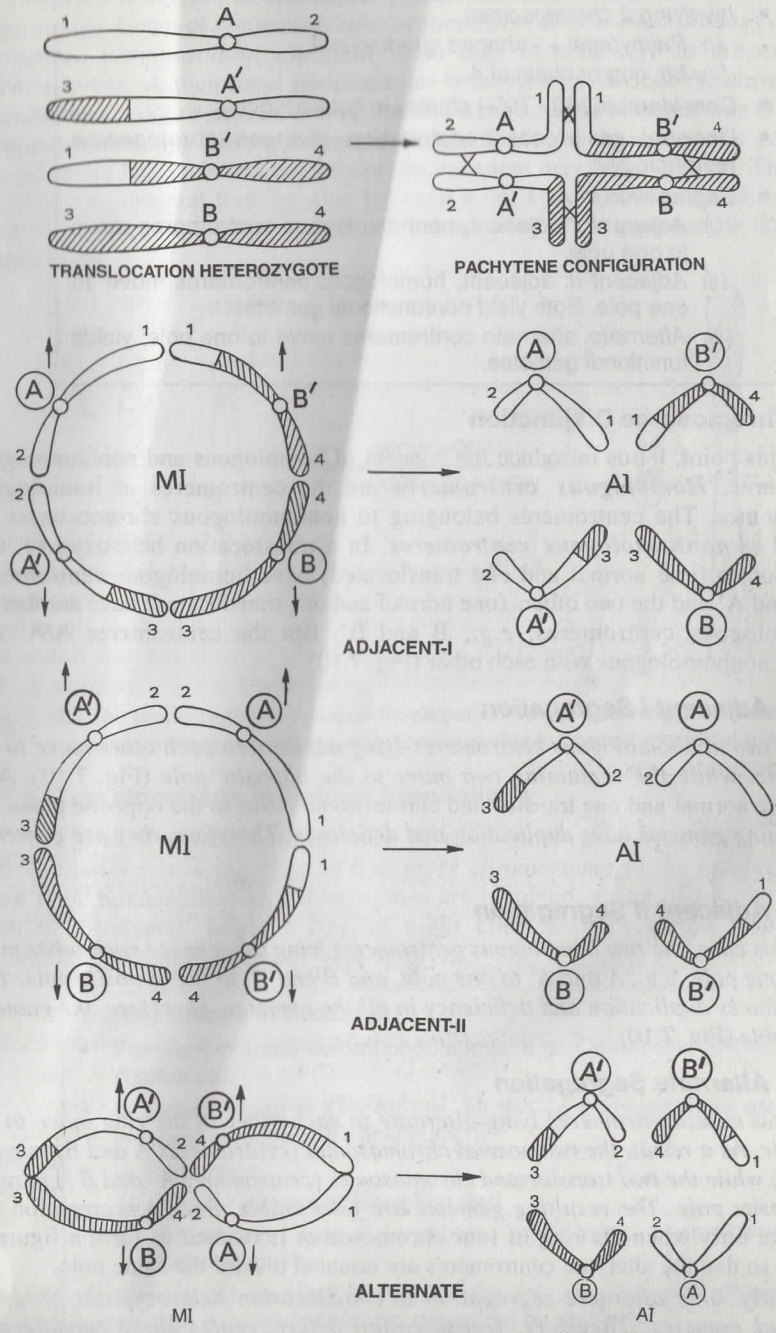


Fig. 7.10. The consequences of adjacent and alternate disjunctions in a translocation heterozygote.

Reciprocal Translocation

- Two non-homologous chromosomes exchange segments.
- *Involving 2 chromosomes* :
  - (*i*) *Pachytene*: + - shaped quadrivalent.
  - (*ii*) *MI*: ring or chain of 4.
- Considerable (60 - 70%) sterility in heterozygotes.
- Unequal reciprocal translocation changes chromosome morphology.
- *Segregation at AI*:
  - (*i*) *Adjacent I*: adjacent, nonhomologous centromeres move to one pole,
  - (*ii*) *Adjacent II*: adjacent, homologous centromeres move to one pole. Both yield nonfunctional gametes.
  - (*iii*) *Alternate*: alternate centromeres move to one pole; yields functional gametes.

7.7.4. Chromosome Disjunction

At this point, let us introduce the concept of homologous and nonhomologous centromeres. ***Homologous centromeres*** are the centromeres of homologous chromosomes. The centromeres belonging to nonhomologous chromosomes are regarded as ***nonhomologous centromeres***. In a translocation heterozygote, two chromosomes (one normal and one translocated) have homologous centromeres, *e.g.*, A and A' and the two others (one normal and one translocated) have another set of homologous centromeres, *e.g.*, B and B'. But the centromeres A/A' and B/B' are nonhomologous with each other (Fig. 7.10).

7.7.4.1. Adjacent I Segregation

The two nonhomologous centromeres lying adjacent to each other move to the same pole, while the remaining two move to the opposite pole (Fig. 7.10). As a result, one normal and one translocated chromosome move to the opposite poles. *All the resulting gametes have duplication and deficiency. Therefore, they are generally inviable.*

7.7.4.2. Adjacent II Segregation

In this case, the two homologous centromeres lying adjacent to each other move to the same pole, e.g., *A and A' to one pole, and B and B' to the opposite pole. This also produces duplication and deficiency in all the gametes. Therefore, the gametes are inviable* (Fig. 7.10).

7.7.4.3. Alternate Segregation

In this case, *centromeres lying alternate to each other in the ring move to the same pole. As a result, the two normal chromosomes (centromeres A and B) move to one pole, while the two translocated chromosomes (centromeres A' and B ') move to the opposite pole. The resulting gametes are fully viable.* Such a segregation can take place only when the ring of four chromosomes is twisted to form a figure of eight (8) so that the alternate centromeres are oriented toward the same pole.

Clearly, *only alternate segregation in translocation heterozygotes produces functional gametes. Therefore, translocation heterozygotes show considerable sterility (generally about 60-70%).*

7.7.5. Change in Chromosome Morphology

Simple translocations and shifts, if large enough, lead to altered chromosome morphology. Unequal reciprocal translocations also alter the morphology of the two concerned chromosomes. They may also lead to a change in the number of chromosomes. A reciprocal translocation between two acrocentric chromosomes may produce a dot or microchromosome and a metacentric chromosome (Fig. 7·11). This phenomenon is known as ***centric fusion.*** Since the regions near a centromere are generally heterochromatic, the dot chromosome may be easily lost. The reverse process, *i.e.*, unequal translocation between a dot and a metacentric chromosome, would produce two acrocentric chromosomes. This process is known as ***dissociation***.

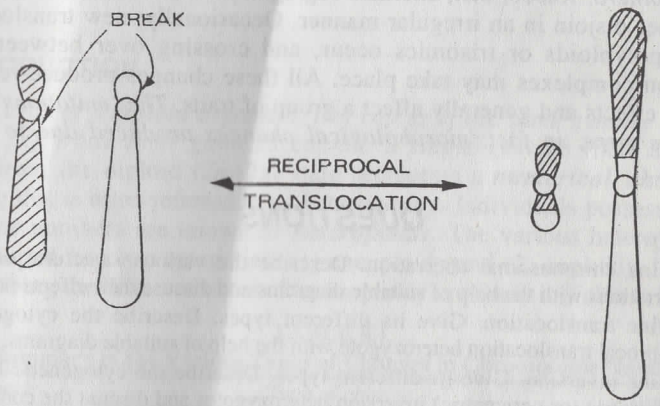


Fig. 7.11. The production of a metacentric chromosome and a microchromosome from two acrocentric chromosomes, and *vice-versa,* due to unequal reciprocal translocation.

7.7.6. Translocations in Natural Populations

Reciprocal translocations may involve more than two nonhomologous chromosomes producing rings of 6 *or more chromosomes in the heterozygotes.* In case three nonhomologous chromosomes are involved, a ring of six chromosomes will be produced, while a ring of eight chromosomes would result if four nonhomologous chromosomes were involved.

Translocations in Evolution

- Frequent in some natural populations, *e.g., Oenothera, Rhoeo, Clarkia* etc.
- In somes races of *Oenothera,* all the 14 chromosomes are involved in reciprocal translocations.
- In *Oenothera*:
  - (*i*) regular alternate disjunction and
  - (*ii*) balanced lethal system maintain permanent heterozygosity.
- Occasional crossing over produces recognisable phenotypic effects, *e.g.,* the 'mutations' described by de Vries.

In most of the *Oenothera* races, two to all of the seven nonhomologous chromosomes are involved in reciprocal translocations. In *Oenothera lamarckiana,* six nonhomologous chromosomes are involved in reciprocal translocations producing a ring of 12 and one bivalent at MI. *On self-pollination (the natural mode of pollination in* Oenothera), O. lamarckiana *produces 50% inviable seeds; the viable seeds are all translocation heterozygotes producing a ring of* 12. *This has been explained on the basis of regular alternate disjunction and a balanced lethal system.* Both zygotic and gametic lethals ensure the production of only translocation heterozygote progeny. This mechanism has been quite successful in *Rhoeo discolor* as well. Rings are maintained in other plant species, *e.g., Clarkia, Paeonia* etc., but they are not enforced through a balanced lethal system.

In *Oenothera, Rhoeo,* etc., alternate segregation is the rule. But sometimes the chromosomes disjoin in an irregular manner. Occasionally, new translocations are produced, polyploids or trisomics occur, and crossing over between different translocation complexes may take place. All these changes produce recognisable phenotypic effects and generally affect a group of traits. *The 'mutations' described by de Vries were, in fact, morphological changes produced due to the above phenomena.*

QUESTIONS

1. Define chromosomal aberration. Describe the various structural chromosomal aberrations with the help of suitable diagrams and discuss their effects on organisms.
2. Define translocation. Give its different types. Describe the cytogenetics of a reciprocal translocation heterozygote with the help of suitable diagrams.
3. Define inversion. Give its different types. Describe the cytogenetic behaviour of paracentric (or pericentric) inversion heterozygotes and discuss the consequences of crossing over within the inversion loop.
4. Describe the various structural chromosomal aberrations, which may alter chromosome morphology and/or number with the help of suitable diagrams.
5. Draw the pachytene configurations when the following are in the heterozygous condition:
   (*i*) deficiency,
   (*ii*) duplication,
   (*iii*) inversion,
   (*iv*) translocation.
6. Write short notes on the following :
   (*i*) duplication,
   (*ii*) deficiency,
   (*iii*) inversion,
   (*iv*) reciprocal translocation,
   (*v*) origin of structural chromosomal aberrations.

8

Chromosomal Aberrations
II. Numerical Aberrations

8.1. INTRODUCTION

In somatic cells of a ***diploid*** *organism, two copies of the same genome are present (i.e.,* $2n = 2x$*), while their gametes contain a single genome (that is,* $n = x$*).* A deviation from the diploid ($2n=2x$) state represents a ***numerical chromosomal aberration;*** this is often referred to as ***heteroploidy.*** Individuals possessing variant chromosome numbers are known as ***heteroploids.*** The various heteroploid states may be grouped into two categories: (1) aneuploidy and (2) euploidy. The various terms describing the different states of heteroploidy are listed in Table 8.1.

TABLE 8.1
A Summary of the Various Types of Changes in Chromosome Number

Term	*Type of change*	*Symbol**
Heteroploid	Change from the $2x$ state.	
A. *Aneuploid*	One or a few chromosomes extra or missing from $2n$.	$2n \pm$ few
Nullisomic	One chromosome pair missing.	$2n - 2$
Monosomic	One chromosome missing.	$2n - 1$
Double monosomic	Two nonhomologous chromosomes (each from a different pair) missing.	$2n - 1 - 1$
Trisomic	One chromosome extra.	$2n + 1$
Double trisomic	Two nonhomologous chromosomes (each from a different pair) extra.	$2n + 1 + 1$
Tetrasomic	One chromosome pair extra.	$2n + 2$
B. *Euploid*	Number of genomes different from two.	
Monoploid	Only one genome present.	x
Haploid	Gametic chromosome complement of the concerned species present.	n
1. *Autopolyploid*	More than two copies of the same genome present	
Autotriploid	Three copies of the same genome.	$3x$**
Autotetraploid	Four copies of the same genome.	$4x$
Autopentaploid	Five copies of the same genome.	$5x$
Autohexaploid	Six copies of the same genome.	$6x$
Autooctaploid	Eight copies of the same genome.	$8x$

2.	*Allopolyploid*	Two or more distinct genomes; generally each genome has two copies.	
	Allotetraploid	Two distinct genomes; each has two copies.	$(2x_1 + 2x_2)^{***}$
	Allohexaploid	Three distinct genomes; each has two copies.	$(2x_1 + 2x_2 + 2x_3)^{***}$
	Allooctaploid	Four distinct genomes; each has two copies.	$(2x_1 + 2x_2 + 2x_3 + 2x_4)^{***}$

* 2n, *somatic chromosome number;* n, *gametic chromosome number of a species, whether diploid or polyploid.*

** x, *the basic chromosome number or genomic number;* 2x, *somatic chromosome number of diploid species.*

*** *In general, this situation occurs; other situations may also occur.*

Aneuploidy

- $2n \pm$ few chromosomes.
- *Discovery*: Bridges in 1916; *X0* and *XXY Drosophila;* types:
  - (*i*) Nullisomy ($2n - 2$).
  - (*ii*) Monosomy ($2n - 1$).
  - (*iii*) Trisomy ($2n + 1$).
  - (*iv*) Tetrasomy ($2n + 2$).
- Weak, reduced vigour and fertility.
- Transmission poor through pollen.
- *Production*: (*i*) spontaneous (*ii*) triploids (best source), (*iii*) *asynaptic* and *desynaptic* plants, (*iv*) translocation hetrozygotes, (*v*) $2n + 2 \times 2n$
- *Uses*: chromosome substitution, locating a gene/linkage group onto a specific chromosome, gene transfer, effects of individual chromosomes, chromosome homoeology.

8.2. ANEUPLOIDY

A loss or gain of one or few chromosomes as compared to the somatic chromosome number of a species is known as ***aneuploidy.*** The aneuploid changes are determined in relation to the somatic ($2n$) chromosome number of the species in question. *Aneuploid changes in chromosome number do not involve the whole genome; they involve only one or a few chromosomes of the genome.*

8.2.1. Terminology

The deficit of a chromosome from the somatic number of a species ($2n$–1) is known as ***monosomic*** condition. The absence of a chromosome pair ($2n$–2) is termed as ***nullisomic*** state. However, if the two missing chromosomes are nonhomologous ($2n$ –1–1), such a condition is known as ***double monosomic*** state. Individuals possessing one extra chromosome ($2n$ + 1) are termed as ***trisomics.*** Those individuals that have an additional chromosome pair ($2n$ + 2) are referred to as ***tetrasomics.*** But if the two additional chromosomes are nonhomologous ($2n$ + 1 + 1), such individuals are known as ***double trisomics.*** These are the most common aneuploid situations.

8.2.2. History

The discovery of aneuploidy dates back to 1916 when Bridges discovered *XO* male and *XXY* female *Drosophila,* which had 7 and 9 chromosomes, respectively, in place of the normal 8. In 1920, Belling demonstrated that the globe mutant of *Datura stramonium* was a trisomic. During 1930's, R.E. Clausen and associates extracted monosomics of tobacco (*N. tabacum*) from interspecific crosses as well as from the *pale sterile* asynaptic mutant. The complete set of 24 ($n = 24$) monosomics in tobacco, the set of 21 monosomics as well as nullisomics in wheat ($n = 21$), the set of 26 monosomics in cotton, and several monosomics and nullisomics in oats (*Avena sativa)* have been developed.

Nullisomy

- $2n - 2$.
- Viable in highly polyploid species, *e.g.,* wheat, oat.
- Weak, reduced size and fertility.
- Characteristic morphological effects for different chromosomes.
- Difficult or unable to be propagated.
- Obtained afresh from progeny of monosomics.
- Limited use in genetic analyses.
- $n - 1$ bivalents during meiosis ; $n - 1$ gametes.

8.2.3. Nullisomy

In ***nullisomy,*** one pair of chromosomes is missing from the somatic chromosome complement. Therefore, it is represented as $2n$–2.

8.2.3.1. Morphological Effects

The absence of one chromosome pair has a strong deleterious effect on the organism. As a result, *nullisomics are viable only in polyploid species, e.g.,* hexaploid or bread wheat (*Triticum aestivum*) and oats (*A. sativa*). But nullisomics do not survive in tobacco (*Nicotiana tabacum* and *N. rustica*), which are allotetraploid. In wheat, *nullisomics are markedly weaker and show reduced size and fertility as compared to normal plants. Nullisomic condition for different chromosomes produces distinct morphological effects.* The morphology of nullisomics for different chromosomes is fairly indicative of the chromosome involved.

8.2.3.2. Origin and Propagation

Many of wheat nullisomics cannot be propagated, but others can be propagated only with difficulty. Therefore, nullisomics are commonly obtained afresh in each generation from monosomic plants. They occur in a low frequency (Ca. 3%) in the selfed progeny of monosomics. Consequently, nullisomics are of relatively little value in practical aneuploid programmes.

8.2.3.3. Cytology

Chromosome pairing in nullisomics is ordinarily normal. At MI, n-1 *bivalents are seen. These bivalents disjoin normally to produce all* n-1 *gametes. Therefore, selfing produces only nullisomic progeny.*

Monosomy

- $2n - 1$.
- *Discovery*: Clausen in 1930's in tobacco.
- Generally not viable in diploids, except maize, tomato etc.
- Variable reduction in vigour and fertility.
- Often distinct morphological effects for different chromosomes.
- During meiosis, $n - 1$ bivalents + 1 univalent.
- At AI the univalent, (*i*) may move to one pole; (*ii*) divide as in motosis; or (*iii*) remain in the centre and would be lost.
- > 50% $n - 1$ gametes.
- Sometimes univalent shift occurs.
- Used in genetic analyses

8.2.4. Monosomy

In ***monosomic condition,*** one chromosome is missing from the somatic chromosome complement. Therefore, this condition is represented as $2n–1$.

8.2.4.1. Morphological Effects

The loss of one whole chromosome creates a major genetic imbalance, which is generally not tolerated by diploid species. Monosomics of polyploid species, *e.g.*, wheat, cotton, oats and tobacco, however, are fully viable, and a complete set of monosomics is available in these crops. In polypoid species like wheat, double monosomics ($2n$ –1–1) as well as triple monosomics ($2n$ –1 – 1 –1) may be obtained.

Some diploid species are exceptional in that their monosomics are viable. Some monosomics of tomato have been recovered. In maize, a complete set of the 10 *monosomics* (n = 10) *is available. Monosomics show a variable reduction in vigour, size, and fertility.* Monosomics of maize and tomato are very weak. Most monosomics of tomato are inviable. But wheat monosomics are comparable to normal plants in vigour and size. Loss of different chromosomes produces distinct morphological effects. But morphology is not always a reliable guide to the type of change at the chromosome level.

8.2.4.2. Cytology

Since one chromosome does not have a pair, it is present as a univalent (unpaired chromosome) at MI. The remaining chromosomes of the complement ordinarily pair normally. *Therefore, n–1 bivalents are produced at MI.*

During AI, the n–1 bivalents disjoin normally. *The single univalent may follow any one of the following three courses.* (1) *It may move to one of the two poles.* (2) *It may lag behind at the metaphase plate and be ultimately lost. Or* (3) *It may, divide into two chromatids* (*as in mitosis*), *which move to the opposite poles* (Fig.8.1). In the last case, the single chromatid of the univalent may follow during AII any one of the first two courses listed above. As a result, *monosomics produce* n–1 *gametics in excess of the expected 50%. In wheat monosomics, about 75% of the female gametes are n–1.*

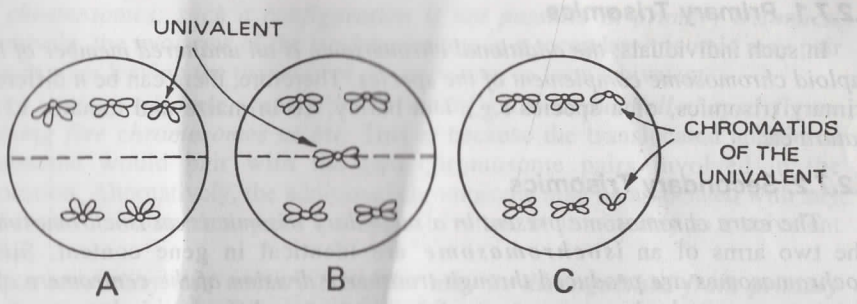


Fig. 8.1. Behaviour at AI of the univalent in a monosomic.

8.2.5. Transmission of the Aneuploid State

Transmission of the aneuploid condition is very poor through pollen grains. This is primarily due to a slower pollen tube growth of aneuploid pollen through the slylar tissue. As a result, normal pollen outcompetes the aneuploid ones.

8.2.6. Univalent Shift

Sometimes, *some progeny plants of monosomics are monosomic for a different chromosome as compared to their monosomic parent. This phenomenon is known as* ***univalent shift.*** Sometimes, monosomics show more than one univalent. Clearly, in such cells some other chromosome (not involved in monosomy) has also remained unpaired. An irregular distribution of the univalents during AI (and may be during AII) in such cells is believed to produce the univalent shift.

Trisomy

- $2n + 1$.
- *Discovery*: Bridges in 1916: *XXY Drosophila* ; Belling in 1920 globe mutant of *Datura.*
- *Types*:
  - (*i*) *Primary*: The extra chromosme is normal member of genome; n different types; 1^{III} or $1^{II} + 1^{I}$
  - (*ii*) *Secondary*: The extra chromosme is an isochromosome; $2n$ different types; 1^{III} or $1^{II} + 1^{I}$ or a ring of 3.
  - (*iii*) *Tertiary*: The extra chromosme is a translocated chromosome; many different types; 1^{III} or $1^{II} + 1^{I}$, or a dumbell of 5 chromosomes.
- Viable in diploids: complete set in maize, tomato, pea etc.
- Variable effect on phenotype; some comparable to $2n$, *e.g.,* maize trisome 5.
- Trivalent would disjoin 2 : 1.
- Univalent behaves an in monosomy.
- < 50% $n + 1$ gametes.
- Selfing gives $2n$ and $2n + 1$ progeny.

8.2.7. Trisomy

When somatic cells of an organism contain three copies of one chromosome, the condition is known as ***trisomy.*** It is denoted by $2n + 1$. Trisomics are grouped into the following three categories: (1) primary, (2) secondary, and (3) tertiary (Fig. 8.2).

8.2.7.1. Primary Trisomics

In such individuals, *the additional chromosome is an unaltered member of the haploid chromosome complement of the species.* Therefore, there can be *n* different primary trisomics, of a species *e.g.*, 7 in barley, 10 in maize and tomato, 12 in *Datura* etc.

8.2.7.2. Secondary Trisomics

The extra chromosome present in a secondary trisomic is an isochromosome. The two arms of an ***isochromosome*** are identical in gene content. Since *isochromosomes are produced through transverse division of the centromere of a chromosome,* each chromosome can give rise to two different isochromosomes. Therefore, 2*n* different types of secondary trisomics are possible in a species.

8.2.7.3. Tertiary Trisomics

Such an individual has a *translocated chromosome in addition to the normal somatic complement.* A large number of different types of tertiary trisomics are possible in a species.

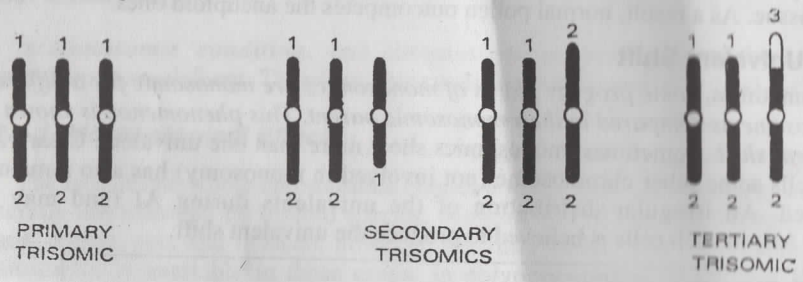


Fig. 8.2. The chromosome constitutions of primary, secondary and tertiary trisomics. Note that the extra chromosome in a secondary trisomic is an isochromosome, while that in a tertiary trisomic it is a translocated chromosome.

8.2.7.4. Genetic Effects

Gain of a complete chromosome causes a major imbalance. But its effects are not as deleterious as those due to the loss of a full chromosome. Therefore, trisomics are viable in diploid species. A full set of *n* primary trisomics is available in maize, tomato, pea, barley, bajra, *Datura,* spinach etc.

Trisomics vary considerably in their vigour and size in relation to the normal disomic plants of the concerned species. Some trisomics are comparable to disomic plants, e.g., trisome 5 in maize. But others show a marked to a great reduction in vigour and size. Often trisomy for different chromosomes produces distinct morphological effects. In some species, morphology is a reasonably reliable guide to the chromosome involved in primary trisomy, *e.g.*, in *Datura.* In some other species, *e.g.*, bajra, however, the different trisomics may not be as distinct morphologically as those in *Datura.*

8.2.7.5. Cytology

In ***primary and secondary trisomics,*** the three homologues may pair with each other to form a trivalent. Alternatively, two of the chromosomes may form a bivalent leaving the additional chromosome as a univalent (Fig. 8.3). The trivalent may show various configurations. *Some PMC's in secondary trisomics would show a ring of*

three chromosomes; such a configuration is not possible in primary trisomics. Alternatively, the two arms of the isochromosome in a secondary trisomic may pair with each other leaving the two normal chromosomes to form a bivalent.

In a ***tertiary trisomic,*** *some PMC's would show a dumbell-shaped figure containing five chromosomes at MI.* This is because the translocated additional chromosome would pair with the two chromosome pairs involved in the translocation. Alternatively, the additional chromosome may be associated with any one of the two chromosome pairs forming a trivalent, or it may stay as a univalent (Fig. 8.3).

Thus ***secondary and tertiary trisomics may be distinguished from primary trisomics due to the presence of a ring of three chromosomes and a dumbell-shaped figure of five chromosomes, respectively, in at least some of their PMC's.***

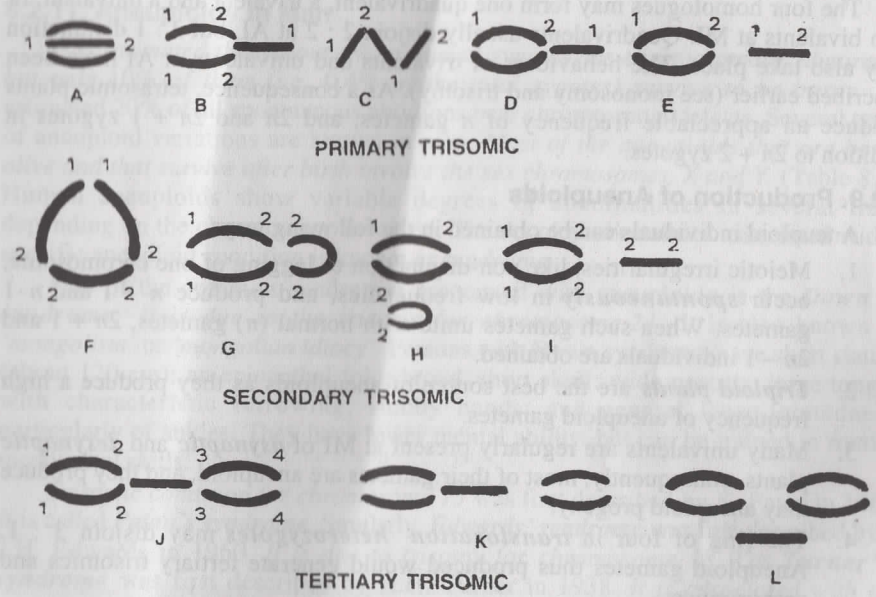


Fig. 8.3. The various metaphase I (MI) configurations of the concerned chromosome in primary, secondary and tertiary trisomics.

During AI, the trivalent (primary and secondary trisomics) would ordinarily show a 2 : 1 disjunction. The univalent would behave irregularly much in the same manner as it does in monosomics. Therefore, the frequency of n + 1 *gametes produced by trisomics is generally appreciably lower than* 50%. Selfing in trisomics produces both normal disomic and trisomic progeny.

Tetrasomy

- $2n + 2$.
- More deleterious than trisomy; survival in polyploids only.
- *At MI* : 1^{IV}, $1^{III} + 1^{I}$ or 2^{II}
- *At AI* :
  - (*i*) Quadrivalent → disjoins 2 : 2 or 3 : 1
  - (*ii*) Trivanlent and univalent → as in trisomy and monosomy.

8.2.8. Tetrasomy

In ***tetrasomic*** $(2n + 2)$ ***individuals,*** somatic cells contain one pair of a chromosome in excess of the normal somatic complement. Thus in tetrasomy, one chromosome of the complement is present in four copies, *i.e.,* it has four homologues.

8.2.8.1. Morphological Effects

In general, gain of a chromosome pair is more deleterious than that of one chromosome. Therefore, *few tetrasomics of diploid species can be recovered and maintained. Tetrasomics are viable in polyploid species,* e.g., *wheat.*

8.2.8.2. Cytology

The four homologues may form one quadrivalent, a trivalent and a univalent, or two bivalents at MI. Quadrivalents usually disjoin 2 : 2 at AI, but 3 : 1 disjunction may also take place. The behaviours of trivalents and univalents at AI have been described earlier (see monosomy and trisomy). As a consequence, tetrasomic plants produce an appreciable frequency of n gametes, and $2n$ and $2n + 1$ zygotes in addition to $2n + 2$ zygotes.

8.2.9. Production of Aneuploids

Aneuploid individuals can be obtained in the following ways.

1. Meiotic irregularities, like non-disjunction or lagging of one chromosome, occur ***spontaneously*** in low frequencies, and produce $n + 1$ and $n-1$ gametes. When such gametes unite with normal (n) gametes, $2n + 1$ and $2n - 1$ individuals are obtained.
2. ***Triploid plants*** are the best source of aneuploids as they produce a high frequency of aneuploid gametes.
3. Many univalents are regularly present at MI of ***asynaptic*** and ***desynaptic*** plants. consequently, most of their gametes are aneuploid, and they produce may aneuploid progeny.
4. The ring of four in ***translocation heterozygotes*** may disjoin 3 : 1. Aneuploid gametes thus produced would generate tertiary trisomics and monosomics.
5. Progeny from a ***cross between tetrasomic*** **(2n + 2)** ***and disomic plants*** show a high frequency of trisomics.

8.2.10. Uses of Aneuploids

Aneuploids are extremely useful in the following several types of genetic studies.

1. They have been used to determine the *phenotypic effects of loss or gain of different chromosomes.*
2. They are used to produce ***chromosome substitution*** lines. Such lines yield information on the effects of different chromosomes of a variety in the same genetic background.
3. They are also used to produce ***alien addition*** and ***alien substitution lines.*** These are useful in gene transfers from one species into another.
4. ***Aneuploid analysis*** *permits the location of a gene as well as of a linkage group onto a specific chromosome.* This is one of the most important applications of aneuploidy.

5. Studies on nullisomic-tetrasomic combinations made it possible to ***establish homoeology*** among the chromosomes of A, B and D genomes of wheat.
6. Multivalent formation in nullisomic 5B of wheat led to the discovery of the *Ph* gene. This gene suppressess pairing between homoeologous chromosomes.
7. *Aneuploids are useful in identifying the chromosomes involved in translocations.*
8. Aneuploids are also useful in preparation of molecular maps.
9. They may be used for obtaining chromosome-specific probes. A ***probe*** is a DNA sequence that is used to detect the presence of the same DNA sequence in a test DNA sample.

8.2.11. Aneuploidy in Man

It is estimated that About 4% of human zygotes are chromosomally abnormal, but only 10% *of them* (i.e., 0.4% *of the total zygotes) survive to be borne.* An estimated 20% of all spontaneous abortions bear chromosomal defects. Several types of aneuploid variations are known in man. *Most of the aneuploids that are borne alive and that survive after birth involve the sex chromosomes, X and Y.* (Table 8.2). Human aneuploids show variable degrees of abnormalities in several traits depending on the chromosome involved. The set of abnormalities associated with a specific aneuploid condition is known as ***syndrome.***

One of the common syndromes associated with aneuploidy is the ***Down's syndrome.*** *It is due to the trisomy for chromosome 21.* It is also known as *'mongolism'* or *'mongolian idiocy'.* Persons with Down syndrome have short stature (about 120 cm); an epicanthal fold; broad, short skull; wide nostrils; large tongue with characteristic furrowing; stubby hands; and general loose-jointedness, particularly of ankles. They have lower mental ability, but can be trained in routine mechanical skills.

Trisomic condition for chromosome 13 was first described by K. Patau in 1960. It is called ***Patau's syndrome.*** Similarly, ***Edwards' syndrome*** was first described by J.H. Edwards in 1960. *It is due to trisomy for chromosome 18.* The ***Turner's syndrome*** was first described by H.H. Turner in 1938. *It is associated with the monosomy for X chromosome.* It has a frequency of about 1 per 2,500 live female births. More than 90% of the *XO* zygotes abort spontaneously. Adults with Turner syndrome are females who have virtually no ovaries and whose secondary sex characters are poorly developed. Such females have short stature, low-set ears, webbed neck and a shield-like chest but generally do not show mental retardation.

Another *trisomic condition for the sex chromosome (XXY) produces* ***Klinefelter's syndrome.*** It was first described by H.F. Klinefelter, in 1942. Such individuals are phenotypically males but with some tendency toward femaleness, particularly in secondary sex characteristics. They may have enlarged breasts, less body hair, small testes and small prostrate glands. *XXY* is the most common karyotype. But Klinefelter patients may be *XXYY, XXXY, XXXYY, XXXXY, XXXXYY and XXXXXY* (Table 8.2). *Patients having more than two X chromosomes show mental retardation as well.*

Rearing and training of abnormal individual is costly. The presence of such individuals in the family causes emotional stress. The expenditure in U.S.A. on the maintenance of persons afflicted by Down syndrome alone is estimated to be around one billion dollars per year.

The incidence of aneuploidy appears to be strongly correlated with the age of mothers. The frequency of trisomy 21 *(Down's syndrome) is about* 1 *in* 1,500 *in mothers of less than* 25 *years, while it is as high as* 1 *in* 100 *in mothers of* 40 *years.*

TABLE 8.2

Aneuploid Variation in Human Beings (*Homo sapiens*). Figures within Parenthesis in the Column Four Represent the Number of Barr Bodies Found in a Cell

Chromosome constitution	*chromosome number*	*Condition*	*Sex chromatin test*
XXY and XXYY	47 48	*Klinefelter's syndrome*: underdeveloped testes; sterility; weak secondary sex characteristics; retarded growth; some mental retardation. Frequency of XXY babies, 1 in 500 live male babies.	+ (1)
XYY	47	Normal male; trisomic; relatively taller; mentally subnormal (IQ, 80-95)	(0)
XO	45	Turner's syndrome: underdeveloped gonads; poorly developed secondary sex characters; sterility; some neck webbing; short stature; low-set ears; shield-like chest; some skeletal abnomality. Frequency, 1 in 2,500 live female births; one in 5,000 adults; 90% of such zygotes abort.	(0)
XXXY and XXXYY	48 49	Extreme Klinefelter's syndrome; severe testicular atrophy; mental retardation; rest symptoms smilar to Klinefelter's syndrome.	+ (2)
XXXXY and XXXXYY	49 50	Extreme Klinefelter's syndrome; malformed genitalia; congenital heart defect; mental retardation.	+ (3)
XXX	47	Triple-X female or metafemale: *phenotypically normal females;* some menstrual upset; some fertility; lack of development of secondary sex chracteristics. Frequency, 1 in 700 live births (female babies).	+ (2)
XXXX	48	*Phenotypically normal females*; mental retardation (metafemales).	+ (3)
XXXXX	49	Metafemale, mental abnormalities.	+ (4)
21, 21, 21	47	*Down's syndrome*: also known as mongolian idiocy; mental retardation; short-body with stubby fingers; swollen tongue, monkey-like skin ridges on extremities; prominent eye-lid folds like those of Mongolian races; one out of every 600 births; average survival 16 years. Frequency, about 1 in 700 among European people.	–
13, 13, 13	47	*Patau's syndrome*: harelip; cleft palate; serious cerebral, occular and cardiovascular defects; survival about 3 months; few have lived upto 5 years. Frequency 1 in 20,000.	—
18, 18, 18	47	*Edward's syndrome*: survival about 6 months; nearly all die within a year; a few have lived upto thirteen years. Mental deficiency; multiple congenital malformation affecting virtually every organ system; frequency, 1 in 8,000.	—

8.3. EUPLOIDY

Euploids have one or more complete genomes, which may be identical with or distinct from each other. In the ***diploid state***, *two copies of the same genome are present in the somatic cells; it is represented as* 2x. *Euploid variations are designated with reference to the diploid* (2x) *state.* These variations may be grouped into the following two broad categories: (1) haploids, including monoploids, and (2) polyploids.

Haploids

- *Haploid* : gametic chromosome number (*n*).
- *Monoploid* : Haploid of a diploid (*x*)
- *Production* : (*i*) spontaneous, (*ii*) delayed pollination, etc., (*iii*) anther culture, (*iv*) interspecific hybridization of barley / wheal with *H bulbosum* and (*v*) ovary culture.
- *n* univalents al MI; random distribution of univalents at AI.
- Highly sterile; few 2*n* progeny.
- Reduced vigour.
- *Uses*: (*i*) homozygous lines, (*ii*) mutant isolation, (*iii*) efficient selection, and (*iv*) all male (XY) progeny.

8.3.1. Monoploidy and Haploidy

Monoploidy *denotes the presence of a single copy of a single genome.* It is represented by *x. Thus monoploids are, in essence, haploids of diploid species.* On the other hand, ***haploidy*** represents the gametic chromosome number of a species irrespective of whether it is a diploid or a ploypoidy species. Haploids are represented by *n*.

8.3.1.1. Production

Haploid plants have to be produced afresh since they cannot be maintained except through clonal propagation. They can be produced in the following ways.

1. Haploids are ***spontaneously produced*** in many plant species in low frequencies. In maize, their frequency varies from 10^{-3} in most of the strains to 10^{-2} in certain strains. *These haploid plants generally develop from unfertilized egg cells* (***parthenogenesis***).
2. The frequency of haploid production may be improved through *delayed pollination, pollination with incompatible pollen, temperature shock* and *X-ray treatment.*

3. *Colchicine treatment of the shoot apical meristem* is reported to induce a small proportion of haploid plants in some species, *e.g.*, sorghum.
4. A much more reliable technique for haploid production is through ***anther*** or ***pollen culture***. Pollen grains of many species, *e.g., Brassica,* rice, tobacco, triticale, etc., when cultured *in vitro* yield haploid plants in appreciably large numbers.
5. *The interspecific cross* **Hordeum bulbosum × H . vulgare** yield haploids of barley. Chromosomes of *H. bulbosum* are selectively eliminated from the developing hybrid embryos. Therefore, most of the plants recovered from this cross are haploids of *H. vulgare.* Similarly, haploid wheat plants are recovered from the cross *H. bulbosum* × *T. aestivum* and from the wheat × maize cross.
6. In many species, culture of unfertilized ovaries yields haploid plants. These plants usually develop from the egg cell, but other cells of the embryo sac may also produce them.

8.3.1.2. Cytology

In haploids, none of the chromosomes has two homologues. Hence *all chromosomes are unpaired during meiosis and are present as* n *univalents at MI. During AI, usually a random sample from the* n *univalents passes to one pole;* the remaining univalents move to the opposite pole. All such gametes are invaiable as they are highly unbalanced. But *sometimes all the* n *chromosomes may pass to the same pole and, thereby, give rise to normal* (n) *viable gametes. Union between such rare and functional male and female gametes produces normal* 2n progeny.

Haploids are highly sterile, but they do produce few 2n *progeny.* Males of some insects, *e.g.,* honey bees, termites, ants etc., are naturally haploid. They are produced from unfertilized eggs. These males are fully fertile since all their *n* chromosomes regularly pass to the same pole during AI.

8.3.1.3. Genetic Effects

In general, haploids (*n*) show a marked reduction in vigour and in the size of various vegetative and reproductive organs as compared to normal disomic (2*n*) plants. But some haploids of potato (2*x*; potato is tetraploid, *i.e.,* 4*x*) are as vigorous and yield as much as the normal disomic (4*x*) strains from which they are derived.

8.3.1.4. Uses

Haploids have the following unique uses in crop improvement.

1. Chromosome doubling of haploids produces ***homozygous disomic*** **(2n)** ***plants in just two years*** as compared to the 6-7 generations of selfing. This greatly reduces the time and labour required for the isolation of inbreds/ purelines. *More than a dozen varieties have been developed in this way* (*most of them in China*).

Polyploidy

- More than two genomes (identical / dissimilar) present.
- *Autopolyploidy*: genomes identical; autotriploids, autotetraploids etc.
- *Allopolyploidy*: genomes dissimilar.
- *Amphidiploids*: Allopolyploids; two copies of each genome; normal bivalent formation at *MI* in natural amphidiploids.

2. They may be useful in the ***isolation of mutants*** because the mutant allele is able to express in the M_1 generation itself.
3. There is some evidence that ***selection at the haploid level may be more efficient*** than that at the diploid level. This is particularly true for autopolyploid species like potato, where breeding is preferred at the haploid level; this approach is called ***analytical breeding.***
4. In some species like asparagus, male plants are more desirable than female plants. Chromosome doubling of asparagus haploids yields YY males. When these YY males are crossed with normal females (XX), all the progeny plants are males (XY). In the normal course (female × normal XY male), only 50% of the progeny plants will be males.

8.3.2. Polyploidy

Presence of more than two genomes in an individual is known as ***ployploidy***. If all the genomes present in an individual are identical, it is called ***autopolyploidy***. But in ***allopolyploidy***, two or more distinct genomes are present. *Naturally occurring allopolyploids ordinarily contain two copies of each of the genomes present. In addition, they show normal bivalent formation like deploids. Therefore, they are known as* ***amphidiploids.***

Autopolyploidy

- *Production*: (*i*) spontaneous (unreduced gametes/somatic chromosome doubling), (*ii*) colchicine, (*iii*) heat/cold/radiation treatments, (*iv*) chemicals like NO_2 etc.
- Gigantism (increased vigour and size usually due to higher water content).
- Larger leaves, flowers, fruits, seeds, pollen grains, stomata.
- Multivalent and univalent formation at MI; irregular disjunction at *AI*.
- *Triploids*: *highly sterile*; seedless watermelon (Japan), sugarbeet (more sugar yield/ unit area) (Europe).
- *Tetraploids*: variable sterility; fertility improved by selection; used as varieties of red clover, rye-gress, berseem (Pusa Giant) etc., ornamentals; Double Steel and Tetra Petkus rye.
- *Evolution*: Some role; potato (4*x*), groundnut (4*x*), coffe (4*x*), lucerne (4*x*) banana (3*x*), sweet potato (6*x*).

8.4. AUTOPOLYPLOIDY

Autopolyploids may have three (***triploid***), four (***tetraploid***), five (***pentaploid***), six (***hexaploid***), seven (***heptaploid***), *eight* (***octaploid***) or more copies of the same genome. Of these, *triploids and tetraploids have been studied in considerable detail due to their importance in crop improvement.*

8.4.1. Production

Autopolyploids are produced through chromosome doubling of a species. This may occur in the following ways.

1. Occasionally, ***unreduced*** **(2n)** ***gametes*** are produced ***spontaneously***. Union between such gametes would yield tetraploid zygotes.

2. Sometimes, ***chromosome doubling*** may occur ***spontaneously*** in the somatic cells giving rise to tetraploid buds.
3. Some of the ***advetitious shoots*** produced following decapitation may be ploypioid. In Solanaceae, 6-36% of such shoots are reported to be tetraploid.
4. ***Heat*** or ***cold treatments*** and ***irradiation of seeds/plants*** with X-rays or gamma-rays may produce ploypioid progeny in a low frequency.
5. A variable frequency of ***plants regenerated from tissue cultures*** of many species, *e.g., Nicotiana, Datura,* rice etc., are ploypioid.
6. ***Some chemicals,*** such as, acenaphthene, 8-hydroxyquinoline and nitrous oxide, induce chromosome doubling.
7. The most efficient method of chromosome doubling is the *treatment of seeds, seedlings or shoot-tips with* ***colchicine***. Colchicine is an alkaloid extracted from the bulbs of *Colchicum autumnale*; it has the chemical formula of $C_{22}H_{25}O_6N$. It interferes with the development of spindle apparatus. As a consequence, the sister chromatids of chromosomes are unable to migrate to the opposite poles during anaphase. Therefore, all the chromatids (= $4n$) are included in the same restitution nucleus leading to chromosome doubling.

Colchicine

- From *Colchicum autumnale* bulbs.
- C_{22} H_{25} O_6 N.
- Unstable in water.
- Prevents spindle organisation.
- *Used as* (*i*) aqueous solution (0·2%), and (*ii*) lanolin paste for 3 - 24 hr (seeds, seedlings) or 2 -3 wks (trees).

Colchicine is generally applied as a 0.2% *aqueous solution or as a mixture with lanolin paste. The duration of treatment may vary from* 3-24 *hr in the case of seeds and seedlings. In case of trees, it may extend upto* 2-3 *weeks.* It is advisable to prepare fresh aqueous solution of colchicine on each occasion since it is unstable in water. The chromosome doubling effect of colchicine was first described by Blakeslee and Nebel independently in 1937.

8.4.2. Genetic Effects

The morphological effect of autoployploidy vary considerably from one species to the other. Some of the features of autopolyploids common to most plant species are summarised below.

1. In many species, autopolyploids show an increase in general vigour and size. This phenomenon is known as '***gigantism'***.
2. In general, leaves of autopolyploids are larger and thicker, and their flowers, fruits and seeds are larger. However, their numbers are relatively smaller as compared to those of normal diploid plants.
3. Cells, pollen grains and stomata of autopolyploids are relatively larger than those of diploids.
4. All autopolyploids show variable sterility.
5. Their growth rate is generally lower than that of normal diploid plants. They are therefore, later in flowering.

6. Usually, autopolyploids show an increase in fresh weight over the diploid strains. However, their *dry weights are either comparable to or, more generally, lower than those of the normal diploid strains.*
7. The optimum level of ploidy may vary considerably from one species to the other.

8.4.3. Cytology

Autopolyploids are characterised by the presence of multivalents at MI. ***Multivalents*** are configurations produced by pairing among more than two homologous chromosomes. In ***triploids,*** the three homologues for each chromosome form either a trivalent, or a bivalent and a univalent. During AI, the disjunction of trivalents and the distribution of univalents is irregular producing a range of aneuploid gametes. As a result, triploids produce a range of aneuploid progeny, *e.g.,* trisomics, double trisomics etc. *Most of the gametes of triploids are unbalanced and inviable.* As a result, they are highly sterile, *e.g.,* banana, watermelons etc. But triploids of some species are highly fertile, *e.g.,* spinach (*Spinacea oleracea*).

In ***autotetraploids***, the four homologues of each chromosome may pair as one quadrivalent, a trivalent and univalent, or two bivalents at MI. The disjunction of trivalents and the behaviour of univalents at AI is usually irregular, while quadrivalents may often dissociate as 2 : 2. As a result, *a variable proportion of the gametes produced by tetraploids are unbalanced and inviable. The fertility of autotetraploids has been considerably improved through selection in several species.*

8.4.4. Role in Evolution

Autopolyploidy has contributed to a limited extent in evolution of plant species. Some of our crops are believed to be autopolyploids, *e.g.,* potato ($4x$), peanut ($4x$), coffee ($4x$), lucerene ($4x$), banana ($3x$) and sweet potato ($6x$). Autotetraploids appear to have been more successful as crops than other forms of autopolyploidy. In addition, many forage grasses and several ornamental species are most likely autopolyploids.

8.4.5. Applications in Crop Improvement

Autopolyploidy has found a limited application in crop improvement. Many autopolyploids have some desirable features. But their usefulness is limited by one or more undesirable characteristics.

Triploid watermelons are produced by crossing tetraploid ($4x$, female) and diploid ($2x$, male) lines. These watermelons produce only rudimentary seeds, which are not objectionable when chewed. But some normal-sized seeds may be formed, satisfactory pollination may be a problem and fruits may be irregular in shape. Triploid watermelons have been commercially grown in Japan.

Triploid sugarbeets produce larger roots and more sugar per unit area than diploids. But seed production of triploid sugarbeet (from $4x \times 2x$ crosses) is difficult. Some $3x$ sugarbeet varieties are grown in Europe.

An ***autotriploid*** clone of tea, TV29, has been released for commercial cultivation in India. TV29 gives higher leaf yield than diploid clones, and its leaf quality is comparable to that of diploid dones.

Autotetraploid varieties of some forage crops have been successful. The most successful examples are, tetraploid red clover and rye-grass. Other examples are

alsike clover (variety, Tora) and berseem (variety, Pusa Giant). Autotetraploid turnip and cabbage are larger in size, but they also have more water content than diploids. Therefore, they are not attractive commercially. Many ornamentals are autotetraploids. They generally have larger flowers and longer flowering duration than diploids. The only *autotetraploid commercial crop varieties* are of rye, *e.g.*, Double Steel and Tetra Petkus.

8.5. ALLOPOLYPLOIDY

Allopolyploids contain two or more distinct genomes, which are derived from different species. Almost all natural allopolyploids have two copies of each of the genomes they contain. They also show regular bivalents at MI as if they were essentially diploid. In view of these facts they are known as ***amphidiploids.***

8.5.1. Origin

Natural allopolyploids most likely originated through chromosome doubling of F_1 hybrids (produced by chance natural hybridization) between two distinct species. ***Chromosome doubling*** may have occurred in somatic cells of the F_1 hybrids due to irregular mitosis. This would give rise to doubled sectors either in the apical meristem or in exillary buds. Sexual progeny from such branches would be amphidiploid. Alternatively, irregular meiosis may lead to the production of ***unreduced gametes.*** Such gametes may unite to produce amphidiploid progeny. Both the pathways are known to occur in experimental populations.

Allopolyploids

- *Production*:
  - (*i*) F_1 produced by interspecific hybridization; chromosome doubling occurs in F_1 spontaneously (unreduced gametes or somatic chromosome doubling) (***in nature***)
  - (*ii*) Interspecific hybridization followed by colchine treatment of F_1 (***synthesis***).
  - (*iii*) Somatic hybridization.

8.5.2. Synthesis

Experimental production of allopolyploids is known as ***synthesis***. This is achieved through chromosome doubling of interspecific hybrids (hybrids between two distinct species) with the help of colchicine. Allopolyploids produced in this way are often known as ***synthetic allopolyploids.***

Thus synthesis of allopolyploids involves two steps: (1) production of F_1 hybrids by crossing two distinct species, and (2) chromosome doubling of the F_1 hybrids thus obtained. For example, *Triticale* is a synthetic allopolyploid. It is usually produced by crossing tetraploid wheats (*Triticum turgidum, AABB)* with rye (*Secale cereale, RR*). The F_1 from this cross has 21 chromosomes (*ABR*) and is completely sterile. It is sterile because there is no pairing among the chromosomes of *A, B* and *R* genomes. Doubling of the chromosome number of this hybrid produces the allohexaploid *Triticale* (*AABBRR*), which is highly fertile (Fig. 8.4).

Allopolyploids may be developed through the technique of ***somatic hybridization.*** In this technique, walls of somatic cells (usually leaf cells) of the concerned species are enzymatically removed. This yields protoplasts enclosed by

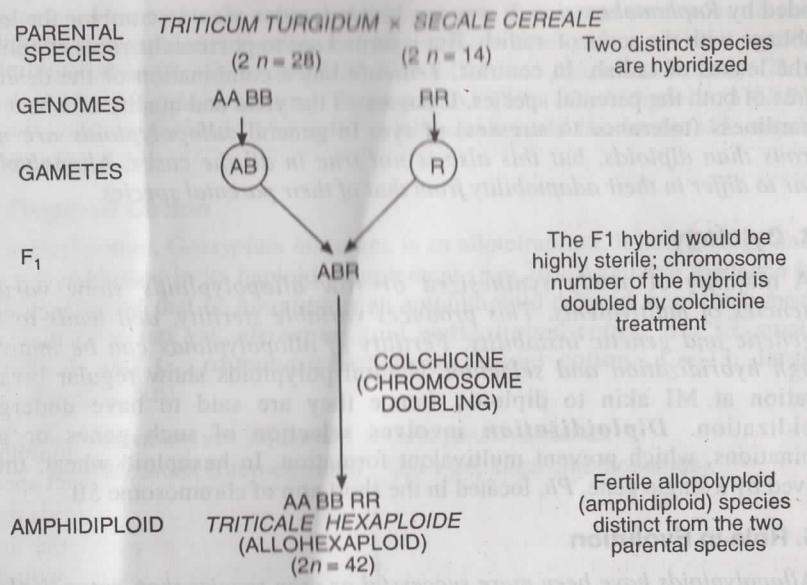


Fig. 8.4. Experimental production (synthesis) of an allopolyploid. The allopolyploid synthesised here is an allohexaploid, the hexaploid triticale (*Triticale haxaploide*).

by plasma lemma only. Protoplasts of the desired species are mixed in equal proportions. They are induced to fuse by a suitable agent, *e.g.*, polyethylene glycol (PEG) treatment. The hybrid protoplasts may be selectively induced to divide on a specifically devised culture medium to yield calli. Later they are induced to regenerate plantlets. This technique has been successfully employed to produce many interspecific hybrids.

8.5.3. Genetic Effects

Allopolyploids generally combine the morphological and physiological features of the parent species. But it is very difficult to predict the combination of the parental characters that would appear in an allopolyploid. An interesting example is

Effects of Allopolyploidy

- More vigorous than parental species.
- Unpredictable combination of parental features.
- Adaptability different from parental species.
- In new allopolyploids, variable frequencies of multivalents and sterility.
- The above improves through hybridization and selection.
- *Diploidization*: due to a single gene *Ph* in wheat (short arm or 5B).
- *Evolution*: About 1/3 angiosperms allopolyploids; crops like wheat, tobacco, oats, sugarcane, *Brassica* spp. etc.
- *Crop Improvement*: (*i*) new crop species, *e.g.*, triticale; (*ii*) as bridging species for gene transfer; (*iii*) creation of genetic variation, *e.g.*, in *B. napus.*

provided by *Raphanobrassica*. It was synthesized with a view to combine the leaves of cabbage with the roots of radish. But it turned out to possess the roots of cabbage and the leaves of radish. In contrast, *Triticale* has a combination of the desirable features of both the parental species. It possesses the yield and quality of wheat with the hardiness (tolerance to stresses) of rye. In general, *allopolyploids are more vigorous than diploids, but this also is not true in all the cases. Allopolyploids appear to differ in their adaptability from that of their parental species.*

8.5.4. Cytology

A majority of *newly synthesized or raw allopolyploids show variable frequencies of multivalents. This produces variable sterility, and leads to both cytogenetic and genetic instability. Fertility of allopolyploids can be improved through hybridization and selection.* Natural polyploids show regular bivalent formation at MI akin to diploids, hence they are said to have undergone diploidization. ***Diploidization*** involves selection of such genes or gene combinations, which prevent multivalent formation. In hexaploid wheat, this is achieved by a single gene, *Ph*, located in the short arm of chromosome 5B.

8.5.5. Role in Evolution

Allopolyploids have been more successful as crop species than autopolyploids. Many of our present-day crops, *e.g.*, wheat, tobacco, cotton, *Brassica* species, sugarcane, oats etc., are allopolyploids. Allopolyploidy has greatly contributed to the evolution of plants. *According to one estimate, about one-third of angiosperms are allopolyploids.* The possible evolutionary history of many allopolyploid crop species has been traced. A brief description of the evolution of some important crops in given below.

8.5.5.1. Origin of Wheat

The common or bread wheat (Triticum aestivum) *is an allohexaploid. It has two copies each of the genomes A, B* and *D*; *its somatic complement is represented as AABBDD.* The sources of *A* and *D* genomes are more or less unanimously accepted as *Triticum monococcum* (*AA*) and *Triticum tauschii* (*DD*), respectively. There is considerable doubt about the source of B genome. According to one hypothesis,

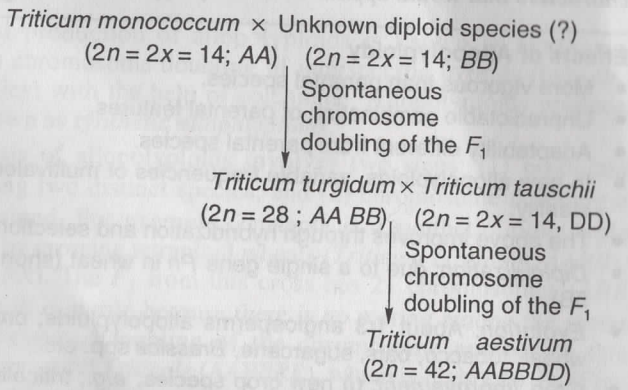


Fig. 8.5. Most probable route of evolution of hexaploid bread wheat (*Triticum aestivum*). The source of *B* genome is still unclear; this is indicated by (?).

Aegilops speltoides may be the source of this genome. But recent evidences do not support this idea. Most likely, the donor of *B* genome is now extinct, and its identity is still not clear. *Most likely, the amphidiploid AABB was produced initially. This gave rise to the tetraploid wheats,* T. turgidum. *This amphidiploid* (*AABB*) *subsequently outcrossed with* T. tauschii *to ultimately yield the hexaploid wheats* (AABBDD) (Fig. 8.5).

8.5.5.2. Origin of Cotton

The upland cotton, *Gossypium hirsutum*, is an allotetraploid. It has 13 large and 13 small chromosomes in its haploid complement (n = 26). Available evidence is interpreted to suggest that *G. hirsutum* is an amphidiploid derived from the hybrid between *G. herbaceum* var. *africanum* (old world diploid cotton, n = 13, small chromosomes) and *G. raimondii* (American diploid cotton, n = 13, large chromosomes) (Fig. 8.6).

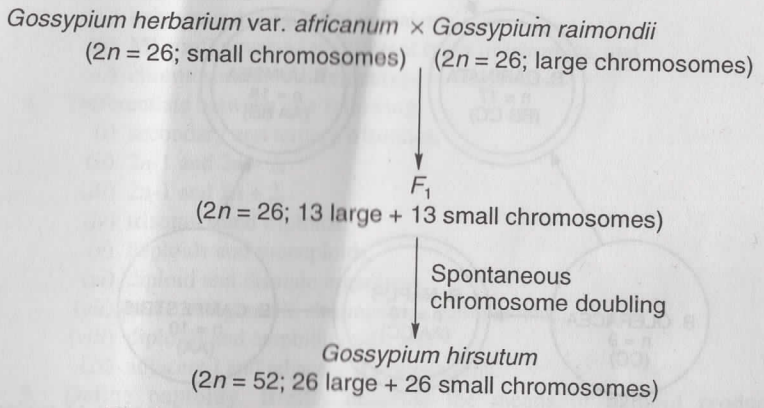


Fig. 8.6. Most likely origin of upland cotton, *Gossypium hirsutum.*

8.5.5.3. Origin of Tobacco

Similarly, *Nicotiana tabacum* and *Nicotiana rustica* are also allotetraploids. Each of them has 24 chromosomes in its haploid complements (2n = 48). Available evidence indicates that *N. tabacum* is an amphidiploid from the cross *N. sylvestrix* × *N. tomentosa.* Each of these two parental species has 12 chromosomes in their haploid complements.

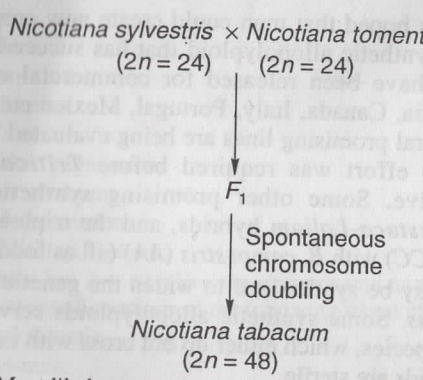


Fig. 8.7. Most likely route of evolution of *Nicotiana tabacum.*

8.5.5.4. *Origin of* Brassica *Species*

Several of *Brassica* species, *e.g., B. juncea, B. napus* and *B. carinata*, are allotetraploids. It is believed that *B. juncea* is an amphidiploid from the cross *B. nigra* × *B. campestris, B. napus* is an amphidiploid from the cross *B. oleracea* × *B. campestris,* and *B. carinata* is an amphidiploid from the cross *B. nigra* × *B. oleracea* (Fig. 8.8).

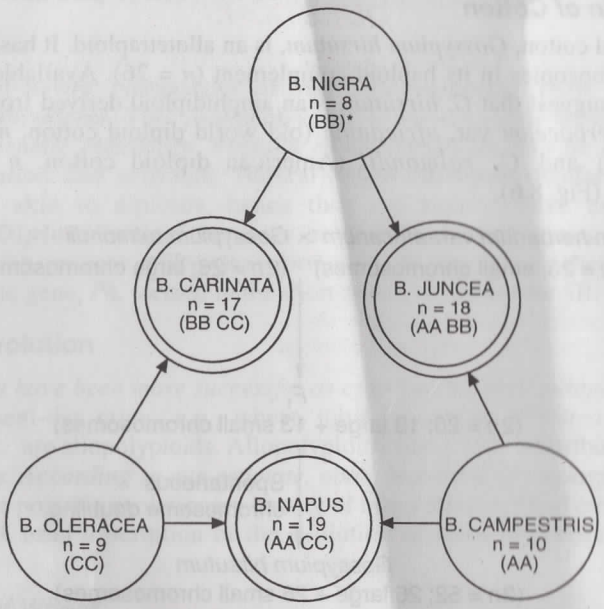


Fig. 8.8. The relationship between diploid and naturally occurring amphidiploid species of *Brassica*. The three diploid species are represented at the three tips of the triangle; their amphidiploids are presented midway between the parental species (and are encircled by two circles).
*Letters within parenthesis denote the genomic symbols for diploid and amphidiploid species.

8.5.6. Applications in Crop Improvement

At one time it was hoped that man could create new crop species at will. But *Triticale* is the only synthetic allopolyploid that has succeeded as a crop species. Varieties of *Titicale* have been released for commercial cultivation in several countries, *e.g.,* Australia, Canada, Italy, Portugal, Mexico etc. In India, TL419 has been released and several promising lines are being evaluated. More than fifty years of intensive research effort was required before *Triticale* varieties became commercially attractive. Some other promising synthetic allopolyploids are *Raphanobrassica, Festuca-Lolium* hybrids, and the triploid (*AAC*) obtained by crossing *B. napus* (*AACC*) with *B. campestris* (*AA*) (all as fodder crops).

Allopolyploids may be synthesized to widen the genetic base of allopolyploid crops, *e.g.,* in *B. napus*. Some synthetic allopolyploids serve as bridging species between two diploid species, which either do not cross with each other or when they do cross, their F_1 hybrids are sterile.

More often synthetic allopolyploids are used as *bridging species* to enable gene tansfer from a related wild species to a crop species. Often interspecific hybrids are sterile. The chromosome numbers of such hybrids is usually doubled and the resulting allopolyploids are ordinarily at least partially fertile. These allopolyploids are now used in a backcross programme with the crop species to allow the transfer of desired genes into the crop.

QUESTIONS

1. Prepare a list of different numerical chromosome aberrations. Define aneuploidy, give its types and discuss their cytological and morphological features.
2. Define euploidy. Describe the cytological and morphological features of autopolyploids (or allopolyploids), the mode of their experimental induction and their role in evolution and crop improvement.
3. Draw the following:
   (*i*) behaviour during AI of univalents,
   (*ii*) MI configurations of different types of trisomics, and
   (*iii*) chromosomes of different types of trisomics.
4. Differentiate between the following:
   (*i*) secondary and tertiary trisomics,
   (*ii*) $2n$-1 and $2n - 2$,
   (*iii*) $2n$-1 and $2n + 2$,
   (*iv*) trisomics and triploids,
   (*v*) haploids and monoploids,
   (*vi*) diploid and disomic organisms,
   (*vii*) basic and gametic chromosome complements,
   (*viii*) diploids and amphidiploids, and
   (*ix*) adjacent I and adjacent II segregations.
5. Define haploidy. Breifly describe the means of halpoid production, their morphological and cytological features and their applications in crop improvement.
6. List the various types of aneuploïds found in man. Briefly describe the feature of some common human aneuploids.
7. Give a list of some of our auto- and allo-polyploid crop species. Briefly describe the mode of origin of some of our allopolyploid crops.
8. Write short motes on the following:
   (*i*) uses of aneuploids,
   (*ii*) bridging species,
   (*iii*) *Triticale,*
   (*iv*) *Oenothera,*
   (*v*) univalent shift,
   (*vi*) diploidization,
   (*vii*) Down's syndrome,
   (*viii*) Klinefelter's syndrome,
   (*ix*) Turner's syndrome,
   (*x*) triploid watermelons,
   (*xi*) synthetic allopolyploids.
9. The bread wheat has $2n$= 42. Give the chromosome numbers, general symbols and MI chromosome configurations of the following wheat plants:
   (*i*) a normal disomic,
   (*ii*) a monosomic,
   (*iii*) a trisomic,

(*iv*) a double trisomic,
(*v*) a double monosomic,
(*vi*) a nullisomic,
(*vii*) nullisomic 5B.

10. In maize, $2n = 20$. Give the chromosome numbers and general symbols for the following maize plants:
(*i*) haploid,
(*ii*) monoploid,
(*iii*) triploid,
(*iv*) tetraploid,
(*v*) hexaploid,
(*vi*) pentaploid,
(*vii*) heptaploid.

9

Laws of Mendel
I. Law of Segregation

9.1. INTRODUCTION

It has been recognised since prehistoric days that progeny of human beings look like human beings, while those of animals like dog and cat are similar to dog and cat, respectively. These observations are expressed as widely quoted proverbs, such as, 'like begets like'. When a child is borne, people look for its resemblance with the parents and close blood relatives. This implies that the development of characters of a baby is somehow related to those of its parents, or of others in its immediate ancestry. Questions relating to the nature and the basis for this relationship have occupied the thoughts of man for centuries. But systematic attempts to seek answers to these questions began only in the eighteenth century when several scientists began studies on plant hybridization. These studies laid the foundation for the investigations of Gregor J. Mendel (1822-1884). They still remain the most brilliant and the most conclusive experiments in genetic analysis supporting the Laws of Mendel.

9.2. CONCLUSIONS OF THE SCIENTISTS BEFORE MENDEL

A number of scientists had worked on plant hybridization during 18th and 19th centuries prior to Mendel. Some of the more notable scientists are Joseph Koelreuter, John Goss, Sargeret, Gaertner, Darwin, Herbert, Lecoq, Vichura and Naudin. Koelreuter conducted extensive studies on hybridization in tobacco between 1760 and 1766. He noted uniformity and heterosis in F_1 (first filial generation; filial = progeny) and appearance of increased variation in F_2.

Gaertner (1722-1850), Naudin (1815-1909), Darwin (1809-1882) and others confirmed the observations and conclusions of Koelreuter. Gaertner used a back-cross programme to convert one species into another. Essentially, he transferred the nucleus of one species into the cytoplasm of another species.

The following important conclusions were available to Mendel from the studies of his predecessors.

1. In F_1 hybrids, some characters are identical to those of one of the two parents, some others are similar to those of the other parent, while some others are intermediate between those of the two parents.
2. Characters of F_1 and F_2 progeny produced by reciprocal crosses are identical. This observation clearly demonstrates that the contributions of male and female parents to the characters of progeny are equal.

3. F_1 progeny from a single cross are uniform in their characters. That is, all the plants in F_1 from a cross are similar to each other. But F_2 generation shows a large variation for different characteristics.
4. In F_2 generation, some plants have characters similar to one parent, while some others are similar to the other parent in their appearance. The appearance of parental forms in F_2 was called ***reversion.*** But a majority of the plants were intermediate in appearance between the two parents.
5. Some plants in F_2 have entirely new character forms.

Conclusions of Predecessors of Mendel

- Uniformity in F_1.
- Heterosis in F_1.
- Increased variation in F_2.
- Of the F_1 characters :
  1. some identical to one parent;
  2. some identical to the other parent;
  3. some are intermediate.
- Equal contribution by both the parents.
- Of F_2 plants :
  1. some similar to one parent;
  2. some to the other parent;
  3. majority intermediate.

9.3. REASONS FOR FAILURE OF MENDEL'S PREDECESSORS

In his 1865 paper, Mendel presented a brilliant analysis of the deficiencies in the experimental approaches of his predecessors. These are summarised below.

1. *These scientists studied the plant as a whole, i.e.*, its total appearance consisting of a large number of characters.
2. Therefore, the *plants could not be classified into few clear-cut classes.* These workers did not attempt an exhaustive classification of the different forms of the characters present in the progeny.
3. The scientists were more concerned with the description of various forms appearing in the progeny. An attempt to determine the frequencies of different character forms in the progeny was not made.
4. In many cases, the data from different generations were not kept accurately and separately.
5. In many cases, a complete control on pollination in the F_1 was lacking.
6. In many studies, the F_1 was an interspecific hybrid exhibiting paratial to considerable sterility.
7. The number of plants studied in F_2 was relatively small.
8. In addition, most of the characters studied by the earlier workers were quantitative in nature.

9.4. GREGOR JOHANN MENDEL

Mendel was born in 1822 near Brunn in Austria, now Brno in Czechoslovakia, in the family of a poor farmer. During his childhood, he had to suffer on account of financial and health problems, which seriously interfered with his education. Unable to continue his studies due to poverty, he joined the St. Augustinian monastery at Brunn in 1843. In 1847 he was made priest. In 1851, he was sent by the monastery authorities to the University of Vienna where he studied physics, mathematics, philosophy etc. Although Mendel was a sincere and hard working student, he did not do very well in his studies, particularly in physics and mathematics.

After completing his studies, he returned to Brunn in 1854 where he was appointed as a substitute science teacher. His performance as a teacher was excellent. In addition, he worked as a priest in the local church. He lived in a house located within the premises of the church. He began to collect pea seeds for his experiments in 1857 from commercial seed growers all over the Europe. He conducted all his experiments within the kitchen garden of his house with the help of his own resources. He carried out seven years of painstaking, sincere, devoted and exhaustive experimentation. He presented his findings before the Natural History Society of Brunn at two of its meetings of February 8 and March 8, 1865. This paper, entitled 'Experiments in Plant Hybridization' was presented in German language. It was published in the annual proceedings of the Society in 1866. This volume of the proceedings was distributed to many libraries of Europe and North America.

Gregor Johann Mendel

- Born in 1822 in Brunn (Austria), now Brno in Czechoslovakia.
- Joined Monastery in 1843.
- In 1847, became priest.
- In 1854 appointed substitute science teacher.
- In 1857, began his famous experiments on pea.
- In 1865, presented his paper before Natural History Society of Brunn.
- In 1866, the paper published in the proceedings of the society.
- In 1900, paper rediscovered by de Vries, Correns Tschermak.

Later, Mendel became more and more involved with the work of monastery. He was made an abbot in 1868. But he still found time to continue his studies on honey bees, some other plants and on climatology. Gregor Mendel died in 1884 at an age of 62 years. In 1900, sixteen years after Mendel's death, three scientists, namely de Vries in Holland, Correns in Germany and Tschermak in Austria, arrived at the same conclusions as those of Mendel. These three scientists, working independently, rediscovered the paper by Mendel. After this rediscovery, there was a spurt of interest in the Mendel's findings. The science of Genetics was thus truly borne.

9.5. SELECTION OF PEA AS AN EXPERIMENTAL MATERIAL

The choice of pea for hybridization by Mendel was based on a deep understanding of the problems of such studies. Pea offered the following advantages as an experimental material.

Selection of Pea

- Self-pollinated.
- Large flowers.
- Easy to grow.
- Needs small space.
- Reproduces in one season.
- Varieties had contrasting forms of several characters.

1. In the available varieties, several characters had two contrasting forms, which were easily distinguishable from each other (Table 9·1). This permitted an easy classification of F_2 and F_3 progeny from various crosses into clear-cut classes.
2. The flower structure of pea ensures self-pollination. This was experimentally verified by Mendel.
3. Pea flowers are relatively large. Therefore, emasculation and pollination of pea flowers is quite easy.
4. The duration of pea crop is of a single season. As a result, every year one generation of pea can be grown.
5. Pea seeds are large and present no problem in germination. Pea plants are relatively easy to grow and each plant occupies only a small space.

In addition to pea, Mendel worked on rajma (*Phaseolus vulgaris* L.) as well. The results from the experiments on rajma were reported along with those on pea in the same paper.

Reasons for Mendel's Success

- Clear understanding of reasons for the failure of earlier workers.
- Study of one character at a time.
- Study of contrasting forms of characters.
- Classification of plants into contrasting groups.
- Accurate record keeping.
- Knowledge of mathematics.
- Formulation and testing of appropriate hypotheses.
- Luck of the brave !

9.6. REASONS FOR MENDEL'S SUCCESS

The following factors together account for Mendel's success.

1. *The most important factor was his ability for an accurate and clear-cut analysis of the reasons for failure of earlier workers.*
2. *At first, Mendel studied the inheritance of only one pair of contrasting characters at a time.* Only after he had formulated the law governing the inheritance of single pairs of contrasting characters, he tried to analyse the inheritance of two and three character pairs.

3. *Mendel selected pea varieties that had clearly different forms of one or more characters.* The difference between the two forms of a pair of contrasting characters was so large and clear-cut that the individuals of a population could be easily and accurately classified as having one or the other contrasting character.
4. *Mendel classified all the plants of a population on the basis of the contrasting characters under study. He kept an accurate record of the numbers of plants (or seeds) in each category for every generation.*
5. Mendel carried out his experiments with great care and elaborateness.
6. His knowledge of mathematics was a definite asset for the interpretation of his findings.
7. *Mendel was able to formulate appropriate hypotheses on the basis of the explanation he offered for his experimental findings. Further, he proceeded to test these hypotheses experimentally to prove the correctness of his explanations.*

But in the final analysis, *Mendel was undoubtedly lucky.* (1) The seven characters selected by Mendel showed qualitative inheritance. Not a single trait was inherited quantitatively. (2) The contrasting forms of each of the seven characters were governed by a single gene. (3) Further, in each case one form was completely dominant over the other. (4) Of the seven characters studied by Mendel, the genes for two were located in one chromosome, while three others were present in another chromosome. *But out of these, only two where close enough to distort the dihybrid* ratio of 9 : 3 : 3 : 1. *Luckily, Mendel did not study this character pair.*

Laws of Mendel

- Law of segregation.
- Law of independent assortment.

9.7. LAWS OF MENDEL

From his studies in pea, Mendel proposed the two basic laws of genetics: (1) law of segregation and (2) law of independent assortment (Chapter 9). These laws provided the foundation on which the science of genetics has developed. *The law of segregation is universally applicable, but the law of independent assortment does not apply to linked genes.*

Mendel also explained the phenomenon of dominance. But dominance is not regarded as a law. More information on dominance is provided in Chapter 11.

9.8. LAW OF SEGREGATION

According to this law, *the two alleles of a gene remain separate and do not contaminate each other in the* F_1 *or the hybrid. At the time of gamete formation in* F_1, *the two alleles separate and pass into different gametes.*

Thus each somatic cell of an individual has two copies (or alleles) of a gene. But each gamete has only a single copy (or allele) of any gene. As a result, a somatic cell may be either pure (containing two identical copies of a gene, *e.g.*, *AA*) or hybrid (containing two dissimilar copies of a gene, *e.g.*, *Aa*) for a gene. But, of necessity, the gametes are always pure for every gene since they have only one copy

of each gene (either *A* or *a*). Therefore, the term '***law of purity of gametes***' is sometimes used to describe the law of segregation.

Segregation for a gene produces two types of gametes (*e.g.,* A *and* a) *in equal frequency.* A random union among these male and female gametes produces a 3 : 1 ratio in the F_2 generation. *The 3 : 1 ratio is the typical monohybrid ratio.* (Here, we have assumed complete dominance.)

The law of segregation was discovered from studies on monohybrid crosses in pea. A ***monohybrid*** is obtained by crossing two varieties or strains, which differ from each other for a single character. For example, Mendel crossed a pea variety having round seeds with another variety possessing wrinkled seeds. The varieties or strains, which differ from one another for one or more characters and are involved in a cross are known as ***parents.*** In this cross, the two parent differed from each other for a single character, *viz.*, seed shape. Similarly, a cross between parents differing for two characters is known as ***dihybrid cross.***

TABLE 9·1
A Description of the Seven Pairs of Contrastings Characters of Pea Studied by Mendel

Character	*Dominant form*	*Recessive form*
Seed shape	Round	Wrinkled
Seed coat colour	Grey	White
Cotyledon colour	Yellow	Green
Pod colour	Green	Yellow
Pod shape	Full	Constricted
Position of flowers	Axial	Terminal
Length of stem	Tall	Dwarf

9.8·1. Inheritance of Seed Shape in Pea

Mendel began his studies on a total of 34 distinct varieties of pea (*Pisum sativum* L.). For two consecutive years, he grew each of these varieties in separate plots to (1) check for their purity and to (2) assess the stability of their characters. Finally, he selected 22 of these varieties for hybridization. Each of these varieties differed from the other with respect to one or more characters. A ***character*** is any morphological, anatomical, biochemical or behavioral feature of an organism, *e.g.*, seed shape, cotyledon colour etc. Mendel selected 7 characters for his study (Table 9·1). Each of these characters had two distinctly different forms; such forms are called ***alternative*** or ***contrasting forms,*** *e.g.,* round and wrinkled seed shape; yellow and green cotyledon colour etc.

To begin with, Mendel crossed varieties differing for one pair of contrasting characters. ***Crossing*** or ***hybridization*** consists of placing live pollen grains from flowers of one parent (***male parent***) onto the stigma of emasculated flowers of the other parent (***female parent***). ***Emasculation*** is the removal of all immature anthers (or even androecium) of a flower. *Emasculation is done in the flowers of the female parent of a cross* to prevent self-pollination. The progeny obtained from crossing (hybridization) are known as ***hybrid*** or ***F_1 generation*** (F_1 = *first filial* or *progeny* generation). Self-pollination in F_1 (in animals, intermating among F_1 individuals) produces F_2, while that in F_2 generates F_3 and so on.

In one of his experiments, Mendel crossed a variety producing round seeds with another having wrinkled seeds. The seeds resulting from hybridization (F_1 seeds)

were all round. The F_2 seeds, *i.e.*, seeds produced by selfing of F_1 plants, were of two kinds: on an average 3 out of every 4 seeds were round, while one out of every 4 seeds was wrinkled. Mendel raised the F_2 plants, allowed them to self-pollinate and studied the shape of seeds produced by each F_2 plant. He compared these findings with the shape of seed that produced the F_2 plant in question. He found that all the plants from wrinkled F_2 seeds produced only wrinkled F_3 seeds. But F_2 plants raised from round seeds were of two types: on an average, 1/3 plants produced only round seeds, while 2/3 plants produced both round and wrinkled seeds in the ratio 3 : 1.

Mendel further noted that the results from reciprocal crosses were identical (Fig. 9.1); therefore, he pooled the data from such crosses. When one parent (parent *A*) is first used as female in a cross with another parent (parent *B*), and then it (parent *A*) is used as male in another cross with the same parent (parent *B*), such two crosses are known as ***reciprocal crosses***. For example, the crosses $A \times B$ and $B \times A$ are reciprocal crosses. *As a rule, the female parent is shown first in a cross,* e.g., *in cross* A × B *the parent* A *is used as female, while in cross* B × A *the parent* B *is used as female.*

TABLE 9·2
A Summary of the F_2 Data Obtained by Mendel in Peas from Different Monohybrid Crosses

Character	*Dominant*	*Recessive*	*Total*	*Ratio*
Seed shape	5,474	1,850	7,324	2.96 : 1
Cotyledon colour	6,022	2,001	8,023	3.01 : 1
Seed coat colour	705	224	929	3.15 : 1
Pod colour	428	152	580	2.82 : 1
Pod shape	882	299	1,181	2.95 : 1
Position of flowers	651	207	858	3.14 :1
Length of stem	787	277	1,064	2.84 : 1

Results obtained from monohybrid crosses for the remaining six characters were identical with those for seed shape (Fig. 9.1, Tables 9.1, 9.2). The following conclusions can be drawn from these findings.

1. *Both male and female parents make equal contribution to the development of characters in the progeny* since the results from reciprocal crosses are identical.
2. In F_1 generation, character of only one of the two parents is expressed; this character is known as ***dominant*** character. Character of the other parent is not expressed in F_1; such a character is referred to as ***recessive.***
3. *In* F_2 *generation, characters of both the parents,* i.e., *both dominant and recessive characters, appear in a definite proportion of 3 dominant : 1 recessive.*
4. *The recessive character appears in F_2 unchanged and identical to that of the parent contributing this trait.* Clearly, the recessive character is not modified in the F_2 generation. Simply its expression is prevented.
5. *The dominant character can be of two types: firstly,* it may be ***pure*** like that in the parent, and *secondly,* it may be ***hybrid*** similar to the F_1 hybrid.

of each gene (either *A* or *a*). Therefore, the term ***'law of purity of gametes'*** is sometimes used to describe the law of segregation.

Segregation for a gene produces two types of gametes (*e.g.*, A *and* a) *in equal frequency.* A random union among these male and female gametes produces a 3 : 1 ratio in the F_2 generation. *The 3 : 1 ratio is the typical monohybrid ratio.* (Here, we have assumed complete dominance.)

The law of segregation was discovered from studies on monohybrid crosses in pea. A ***monohybrid*** is obtained by crossing two varieties or strains, which differ from each other for a single character. For example, Mendel crossed a pea variety having round seeds with another variety possessing wrinkled seeds. The varieties or strains, which differ from one another for one or more characters and are involved in a cross are known as ***parents.*** In this cross, the two parent differed from each other for a single character, *viz.*, seed shape. Similarly, a cross between parents differing for two characters is known as ***dihybrid cross.***

TABLE 9·1
A Description of the Seven Pairs of Contrastings Characters of Pea Studied by Mendel

Character	*Dominant form*	*Recessive form*
Seed shape	Round	Wrinkled
Seed coat colour	Grey	White
Cotyledon colour	Yellow	Green
Pod colour	Green	Yellow
Pod shape	Full	Constricted
Position of flowers	Axial	Terminal
Length of stem	Tall	Dwarf

9.8·1. Inheritance of Seed Shape in Pea

Mendel began his studies on a total of 34 distinct varieties of pea (*Pisum sativum* L.). For two consecutive years, he grew each of these varieties in separate plots to (1) check for their purity and to (2) assess the stability of their characters. Finally, he selected 22 of these varieties for hybridization. Each of these varieties differed from the other with respect to one or more characters. A ***character*** is any morphological, anatomical, biochemical or behavioral feature of an organism, *e.g.*, seed shape, cotyledon colour etc. Mendel selected 7 characters for his study (Table 9·1). Each of these characters had two distinctly different forms; such forms are called ***alternative*** or ***contrasting forms,*** *e.g.*, round and wrinkled seed shape; yellow and green cotyledon colour etc.

To begin with, Mendel crossed varieties differing for one pair of contrasting characters. ***Crossing*** or ***hybridization*** consists of placing live pollen grains from flowers of one parent (***male parent***) onto the stigma of emasculated flowers of the other parent (***female parent***). ***Emasculation*** is the removal of all immature anthers (or even androecium) of a flower. *Emasculation is done in the flowers of the female parent of a cross* to prevent self-pollination. The progeny obtained from crossing (hybridization) are known as ***hybrid*** or ***F_1 generation*** (F_1 = *first filial* or *progeny* generation). Self-pollination in F_1 (in animals, intermating among F_1 individuals) produces F_2, while that in F_2 generates F_3 and so on.

In one of his experiments, Mendel crossed a variety producing round seeds with another having wrinkled seeds. The seeds resulting from hybridization (F_1 seeds)

were all round. The F_2 seeds, *i.e.*, seeds produced by selfing of F_1 plants, were of two kinds: on an average 3 out of every 4 seeds were round, while one out of every 4 seeds was wrinkled. Mendel raised the F_2 plants, allowed them to self-pollinate and studied the shape of seeds produced by each F_2 plant. He compared these findings with the shape of seed that produced the F_2 plant in question. He found that all the plants from wrinkled F_2 seeds produced only wrinkled F_3 seeds. But F_2 plants raised from round seeds were of two types: on an average, 1/3 plants produced only round seeds, while 2/3 plants produced both round and wrinkled seeds in the ratio 3 : 1.

Mendel further noted that the results from reciprocal crosses were identical (Fig. 9.1); therefore, he pooled the data from such crosses. When one parent (parent *A*) is first used as female in a cross with another parent (parent *B*), and then it (parent *A*) is used as male in another cross with the same parent (parent *B*), such two crosses are known as ***reciprocal crosses***. For example, the crosses $A \times B$ and $B \times A$ are reciprocal crosses. *As a rule, the female parent is shown first in a cross,* e.g., *in cross* A × B *the parent* A *is used as female, while in cross* B × A *the parent* B *is used as female.*

TABLE 9·2
A Summary of the F_2 Data Obtained by Mendel in Peas from Different Monohybrid Crosses

Character	*Dominant*	*Recessive*	*Total*	*Ratio*
Seed shape	5,474	1,850	7,324	2.96 : 1
Cotyledon colour	6,022	2,001	8,023	3.01 : 1
Seed coat colour	705	224	929	3.15 : 1
Pod colour	428	152	580	2.82 : 1
Pod shape	882	299	1,181	2.95 : 1
Position of flowers	651	207	858	3.14 :1
Length of stem	787	277	1,064	2.84 : 1

Results obtained from monohybrid crosses for the remaining six characters were identical with those for seed shape (Fig. 9.1, Tables 9.1, 9.2). The following conclusions can be drawn from these findings.

1. *Both male and female parents make equal contribution to the development of characters in the progeny* since the results from reciprocal crosses are identical.
2. In F_1 generation, character of only one of the two parents is expressed; this character is known as ***dominant*** character. Character of the other parent is not expressed in F_1; such a character is referred to as ***recessive.***
3. *In* F_2 *generation, characters of both the parents,* i.e., *both dominant and recessive characters, appear in a definite proportion of 3 dominant : 1 recessive.*
4. *The recessive character appears in F_2 unchanged and identical to that of the parent contributing this trait.* Clearly, the recessive character is not modified in the F_2 generation. Simply its expression is prevented.
5. *The dominant character can be of two types: firstly,* it may be ***pure*** like that in the parent, and *secondly,* it may be ***hybrid*** similar to the F_1 hybrid.

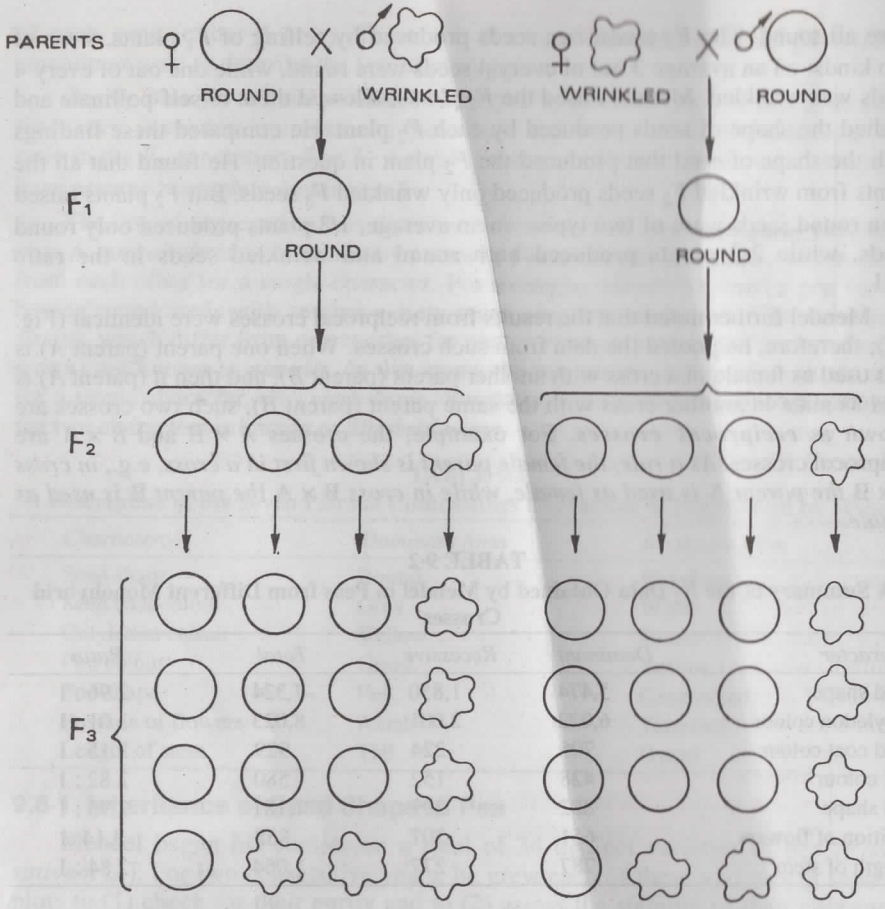


Fig. 9.1. Diagramatic representation of the results obtained by Mendel from reciprocal crosses between round and wrinkled seeded pea plants.

6. In F_2, 1/3 of the individuals with the dominant character are pure, while 2/3 of them are hybrid.

9.8·2. Genetic Explanation

On the basis of these findings, Mendel proposed the first fundamental principle of genetics, the ***law of segregation.*** According to this law, *the two alleles of a gene stay together in the hybrid without contaminating each other, and at the time of gamete formation they separate and pass into different gametes.*

The two important consequences of this law are: (*i*) *each gamete contains only one allele of a gene,* and (*ii*) *a 3 : 1 ratio is obtained in the* F_2 *generation of a monohybrid cross.* Therefore, the 3 : 1 ratio is commonly known as the ***monohybrid ratio.***

The law of segregation can be explained more clearly by making the following assumptions. (1) A character is produced by a specific ***gene.*** (2) Each gene has two

The Law of Segregation
- The first law of Mendel.
- When two alleles of a gene are brought together into an F_1 hybrid, they:
  1. remain separate;
  2. do not contaminate each other;
  3. pass into separate gametes.

Consequences
1. Gametes produced in 1 : 1 ratio by F_1.
2. In F_2, 3 : 1 phenotypic ratio and 1 : 2 : 1 genotypic ratio.
3. In test cross, 1 : 1 phenotypic and genotypic ratios.

alternative forms; these forms are known as ***alleles.*** (3) The two alleles of a gene govern the contrasting forms of the character produced by the gene.

A gene can be represented by a symbol derived from the name of the character it governs. The gene controlling seed shape in pea may be represented by the letter *w* taken from *w*rinkled (seed shape). The symbol for a gene may have a single letter (*e.g., w*) or up to, usually, 3 letters (*e.g., rht* for *r*educed *h*eigh*t* in wheat). *The allele producing the recessive form of the character is represented by small or lower case letters. The symbol for the allele producing the dominant form of the character is the same as that for the recessive allele, but the first letter of this symbol is capital or in upper case.* For example, round seed shape is dominant and is assigned *W* as the symbol, while wrinkled seed shape is recessive and is denoted by *w*. *Thus the two alleles of a gene are represented by the same gene symbol and are differentiated from each other by their first letter being in lower or upper case.*

Essential Features of Inheritance
- Each character is governed by a specific gene.
- Each gene has two alleles.
- Alleles of a gene govern the alternative forms of the same trait.
- Each somatic cell has two copies of each gene.
- Each gamete has only one copy of each gene.
- Both male and female parents contribute equally to the progeny.

(4) *Each somatic cell of an organism has two copies of each gene.* An individual may have two copies of the same allele, *e.g., WW* or *ww*. Such an organism is called ***homozygous*** and the condition is known as ***homozygosity.*** But when an individual has two different alleles of the same gene, *e.g., Ww*, it is called ***heterozygous,*** and the condition is termed as ***heterozygosity***. Mendel used the terms ***pure*** and ***hybrid*** for homozygous and heterozygous, respectively.

Let us now consider definitions of some essential terms. When an allele produces the character governed by it. it is said to ***express*** itself. The allele that expresses itself in the heterozygous state is known as ***dominant.*** But the allele, which is unable to express itself in the heterozygous state is called ***recessive.*** The phenomenon of only one of the two alleles of a gene present in the heterozygotes being able to express itself is known as ***dominance.***

The appearance of an individual with respect to a specific character (or all of its characters) is referred to as ***phenotype.*** But the kinds of alleles of a specific gene (or all the genes) carried by an individual is known as its ***genotype*** or ***genetic constitution.*** A round pea seed may be either homozygous (*WW*) or heterozygous

(*Ww*) genotypically. But in both the conditions it will have the same phenotype (appearance), *i.e.*, round.

The genotype of individuals with a dominant character cannot be known from their phenotype. It can be known only after studying their progeny (either from selfing or from a testcross). On the other hand, wrinkled pea seeds can only have the genotype *ww*. Therefore, *the genotypes of individuals possessing recessive traits can be inferred from their phenotypes themselves.*

A *zygote is produced by the fusion of one male and one female gamete. The contributions of male and female gametes are equal.* Therefore, the gametes should have only one copy of a gene, while the zygote has two.

In Mendel's studies, the round and wrinkled parents were homozygous. The genotypes of round and wrinkled parents, therefore, may be written as *WW* and *ww*. Each of these parents would produce only one type of gamete; round parent would produce *W*, while the wrinkled parent would produce *w* gametes. Fusion of these parental gametes will produce *Ww* zygotes (F_1). Seeds developed from these zygotes will be round in appearance since *W* is dominant over *w* (Fig. 9.2).

The appearance of 3 : 1 ratio in F_2 may be explained as follows. At the time of

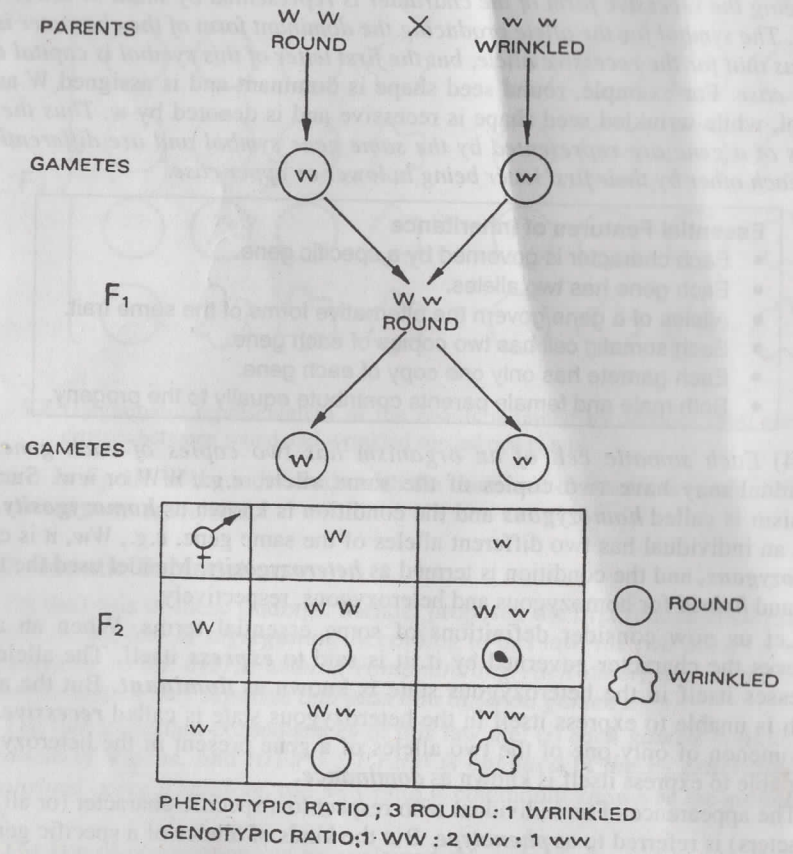


Fig. 9.2. Explanation of the inheritance of seed shape in pea using appropriate gene symbols. (*Commonly referred to as the law of segregation.*)

gamete formation the two alleles, *W* and *w*, present in the F_1 hybrid (*Ww*) separate and go into different gametes. *This is essential because each gamete contains only one copy of a gene.* The hybrid would thus produce two types (*W* and *w*) of male and female gametes in equal frequency, *i.e.*, half of the gametes (of each sex) produced by the hybrid will be *W* and the remaining half will be *w*. A random union of the male gametes with the female gametes of the hybrid would yield four types of zygotes. ***Random union*** of gametes implies that a *W* male gamete has equal chance of uniting with *W* and *w* female gametes. The same applies to *w* male gametes. The four kinds of zygotes produced in F_2 would be *WW*, *Ww*, *Ww* and *ww*; their frequencies will be equal. The first three zygotes would develop into round seeds due to the presence of *W* allele, while the fourth (*ww*) would produce wrinkled seed (Fig. 9.2). This would give rise to the 3 : 1 ratio, which was obtained by Mendel.

If this explanation of the Mendel's observations were correct, (1) the round F_2 seeds should be of two types: *WW* (homozygous) and *Ww* (heterozygous), and (2) the ratio between *WW* and *Ww* F_2 seeds should be 1 : 2. The selfed progeny (F_3) from the *WW* seeds would be all *WW* and round, while *Ww* F_2 seeds would produce in F_3 both round and wrinkled seeds in the ratio 3 : 1. This is precisely what Mendel found in the F_3 generation.

The essential features of this explanation of the 3 : 1 ratio in F_2 are as follows :

1. The presence of two copies of each gene in somatic cells.
2. A lack of contamination or modification of each other by the two alleles of a gene during their stay together in the same cells of F_1 hybrids.
3. Separation of the two copies of a gene at the time of gamete formation.
4. As a result, only one copy of each gene is present in the gametes.
5. The production of two types of gametes, with respect to the heterozygous gene, by the F_1 in equal frequencies.
6. A random union between the male and female gametes.

The concept of noncontamination of each other by the two alleles of a gene is the essence of ***particulate inheritance***. This requires the genes to be like solid particles, so that the two alleles stay separately in F_1. As a result, *the two alleles present in the* F_1 *are able to separate and pass into different gametes in their original form producing two different types of gametes in equal frequencies; this separation of alleles is known as* ***segregation***. A random union between male and female gametes produces the typical 3 : 1 ratio, which is the typical ***monohybrid ratio***. This ratio is observed in the F_2 generation of almost all crosses segregating for a single character.

9.8.3. Phenotypic and Genotypic Ratios

When individuals present in an F_2 (or any other) generation are classified according to their phenotype, we obtain a ***phenotypic ratio.*** For example, the F_2 seeds in Fig. 9.2 yield a 3 round : 1 wrinkled ratio when classified on the basis of seed shape. But when individual, of a F_2 (or any other generation) are classified on the basis of their genotype, this yields a ***genotypic ratio.*** The F_2 seeds in Fig. 9.2 yield a 1 *WW* : 2 *Ww* : 1 *ww* ratio when classified on the basis of their genotypes. Thus we obtain a 3 : 1 phenotypic ratio and a 1 : 2 : 1 genotypic ratio in the F_2 generation of a monohybrid cross.

Segregation: Assumptions
- F_1 produces gametes in 1 : 1 ratio.
- All the gametes survive and function equally.
- Random union of gametes.
- Equal survival of zygotes.

9.8.4. Assumptions Involved in Segregation as an Explanation for the 3 : 1 Ratio

The explanation of 3 : 1 ratio according to the law of segregation is based on the following assumptions: (*i*) the production of two types of gametes in equal frequencies in F_1, (*ii*) equal survival or function of the different gametes, (*iii*) random union of male and female gametes, and (*iv*) equal survival of different zygotes thus produced.

9.9. EVIDENCE FOR EQUAL FREQUENCY OF GAMETES

The evidence for equal frequency of the two types of gametes produced by segregation of a single gene is both indirect as well as direct. ***Indirect evidence*** is based on the frequencies of different types of zygotes produced in a testcross. The ***direct evidence***, on the other hand, comes from a direct classification of gametes and scoring the frequencies of the two classes.

9.9.1. Testcross

A cross between the F_1 hybrid and a strain having the recessive form of the character is known as ***testcross*** (Fig. 9·3). *The purpose of a testcross is to obtain evidence that segregation for the alleles of a single gene in the* F_1 *hybrid produces two types of gametes in equal frequencies.*

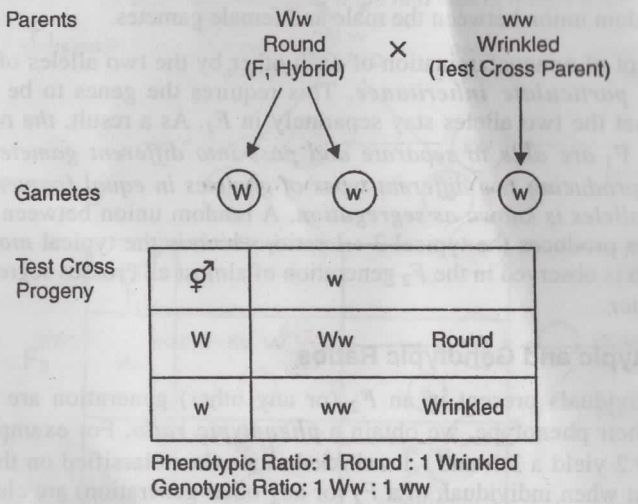


Fig. 9.3. Evidence from testcross data that the F_1 hybrid (*Ww*) produces the two gametes, *W* and *w*, in equal frequencies.

According to the law of segregation, the F_1 (*Ww*) from a cross between round (*WW*) and wrinkled (*ww*) seeded pea varieties would produce gametes *W* and *w* in

equal frequencies. The strain with the recessive trait (*ww*, wrinkled seed) used in the testcross, would produce only one type of gametes, *i.e.*, *w*. Union of the gametes from F_1 with those from the testcross parent, therefore, is expected to produce two types of seeds: *Ww* (round) and *ww* (wrinkled). The frequencies of these two types of seeds are expected to be equal (Fig. 9·3).

Mendel devised and made suitable test crosses in his study. He found that *the ratio in testcross progeny was always 1 dominant (round) : 1 recessive (wrinkled). This indicated that the* F_1 *gametes carrying dominant and recessive alleles were in equal frequency.*

Test Cross
- F_1 crossed with recessive homozygote.
- Phenotypic ratio 1 : 1.
- *Conclusion* : F_1 produces two types of gametes in 1 : 1 ratio.
- *Confirmation :*
  1. 1 : 1 ratio for waxy : non-waxy pollen.
  2. 1 : 1 ratio for black : albino ascospores in *Neurospora.*

9.9.2. Segregation for *waxy* Gene in Maize

In general, pollen grains of plants, particularly cereals and other monocots, stain deep blue with an iodine solution. This staining reaction is the property of starch grains stored in mature pollen grains. A recessive gene in maize, and some other cereals like barley, modifies starch grains in such way that they produce brown colour (in place of the deep blue) on reaction with iodine. This gene is known as *waxy* and is assigned the gene symbol *wx*. Its dominant allele *Wx* governs the production of normal starch grains that stain deep blue with iodine.

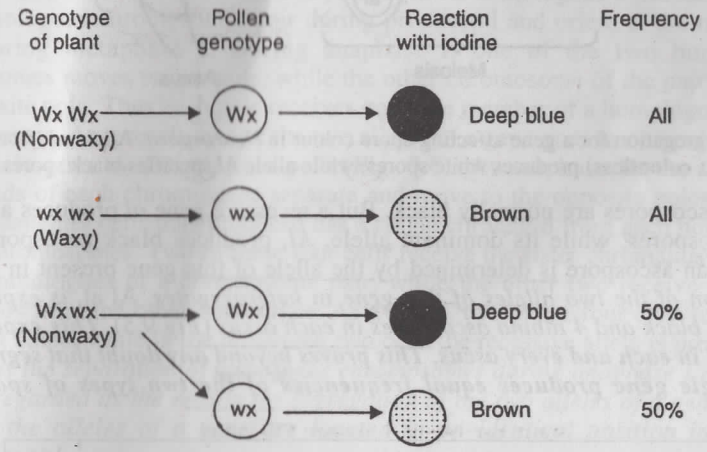


Fig. 9.4. Segregation for the two alleles of the *waxy* gene affecting starch in maize.

The kind of starch grains present in a pollen grain depends on the allele of *waxy* gene present in it. Consequently, pollen grains carrying the *nonwaxy* (*Wx*) allele stain deep blue, while those carrying the *waxy* (*wx*) allele stain brown with iodine (Fig 9.4). This permits a direct classification and counting of the pollen grains from

a *Wx wx* heterozygote for the presence of *Wx* and *wx* alleles in them. This is done by placing a fresh sample of mature pollen grains from a *Wx wx* heterozygote on a glass slide. A drop of an iodine solution is added to the pollen grains, which are then covered with a coverglass and viewed under low power (10x or 40x) of a light microscope. The pollen grains from *Wx wx* heterozygotes will be of two types: *Wx* having deep blue stain and *wx* with brown colour. The total numbers of deep blue and brown coloured pollen grains are counted in several microscopic fields to determine their frequency. The deep blue and brown coloured pollen grains are expected to be in 1 : 1 ratio. Experimentally determined frequencies of these two types agree well with the expected 1 : 1 ratio. This provides a direct evidence that *Wx* and *wx* pollen grains are produced in equal frequencies by *Wx wx* heterozygotes.

9.9.3. Segregation for Spore Colour in *Neurospora*

Ascospore production in *Neurospora crassa* involves the two divisions of meiosis, followed by one mitotic division. As a result, eight haploid nuclei are produced from each diploid zygote nucleus. These nuclei develop into eight haploid ascosproes. The eight ascospores derived from one zygote are enclosed in a membranous structure, which along with the spores is known as ***ascus*** (Fig. 9.5).

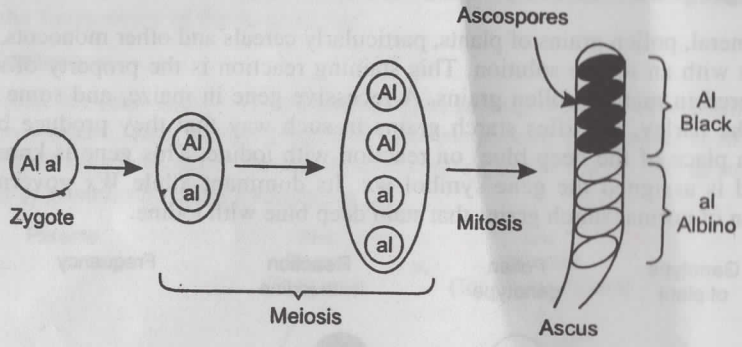


Fig. 9.5. Segregation for a gene affecting spore colour in *Neurospora*. Allele *al* (from *albino, i.e.*, colourless) produces white spores, while allele *Al* specifies black spores.

The ascospores are normally black. But a recessive gene *al* produces albino or white ascospores, while its dominant allele, *Al*, produces black ascospores. The colour of an ascospore is determined by the allele of this gene present in it. *Thus segregation of the two alleles of this gene in heterozygotes,* Al al, *is expected to produce 4 black and 4 albino ascospores in each ascus (Fig 9.5).* ***This expectation is fulfilled in each and every ascus. This proves beyond any doubt that segregation for a single gene produces equal frequencies of the two types of spores (or gametes).***

Segregation : Physical Basis

- The two alleles of a gene are located in homologous chromosomes.
- Separation of homologous chromosomes at Al results in segregation.

9.10. PHYSICAL BASIS OF SEGREGATION

The phenomenon of segregation can be easily explained on the basis of the behaviour of homologus chromosomes during meiosis. As a consequence of segregation, the two alleles of a gene separate and go into different gametes. Similarly, the two members of a homologous chromosome pair during prophase I, and separate at anaphase I of meiosis and move to the opposite poles of a cell. The two chromosomes of a homologous pair are identical in their gene contents, *i.e.*, contain exactly the same set and sequence of genes. Thus one copy of a gene is present in one chromosome of a homologous pair, while the other copy of the gene is present in the second member of the pair. *The two alleles of a gone occupy the same location in the homologous chromosomes.*

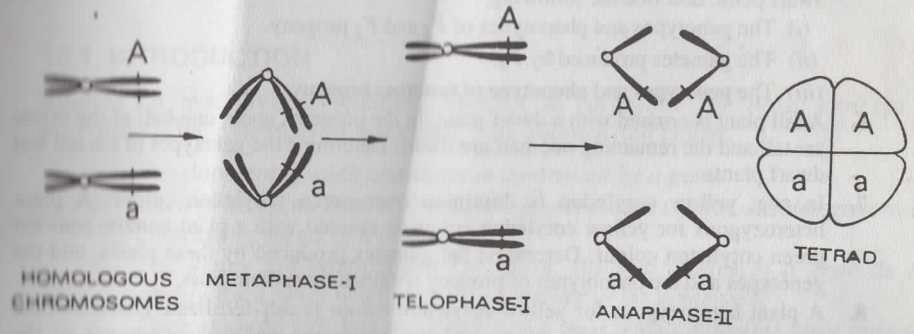


Fig. 9.6. Segregation of a pair of alleles (*Aa*) as a result of movement of the homologous chromosomes and sister chromatids during meosis. In this case, the first meiotic division is the reduction division.

In F_1 hybrid or heterozygote, one of the two alleles of a gene is located in one chromosome, while the other allele is present in its homologous chromosome. The two homologous chromosomes pair during prophase I and orient at the metaphase plate during metaphase I. During anaphase I, one of the two homologous chromosomes moves to one pole, while the other chromosome of the pair moves to the opposite pole. Thus each pole receives only one member of a homologous pair of chromosomes. As a result, one of the two alleles of a gene goes to one pole, and the other allele goes to the opposite pole (Fig. 9.6). At anaphase II, the two sister chromatids of each chromosome separate and move to the opposite poles. Each of the four daughter cells thus produced has a single chromatid from each homologous pair of chromosomes. Two of these four cells receive the sister chromatids from one of the homologues of a chromosome pair, while the other two receive the sister chromatids from the other homologue. As a result, two of the four cells receive the dominant allele, while in the remaining two cells the recessive allele is present (Fig. 9.6). *Thus the separation of homologus chromosomes during anaphase I of meiosis may be regarded as the reason for segregation of the two alleles of a gene. This is because the alleles of a gene are located in an identical position in the two homologous chromosome.*

QUESTIONS

1. Describe the conclusions of workers on plant hybridization before Mendel and discuss the reasons for their failure. Discuss the reasons for Mendel's success.

2. Explain the law of segregation and its consequences with the help of a suitable example.
3. List the assumptions necessary for the explanation of the 3 : 1 ratio in F_2 on the basis of segregation. Discuss the direct and indirect evidences to show that segregation for a gene gives rise to two types of gametes in equal frequencies.
4. Write short notes on the following:
   (*i*) physical basis of segregation,
   (*ii*) dominance,
   (*iii*) Mendel,
   (*iv*) choice of pea by Mendel.
5. In pea, tall plant is dominant over dwarf plant. If a homozygous tall is crossed with a dwarf plant, describe the following.
   (*i*) The genotypes and phenotypes of F_1 and F_2 progeny.
   (*ii*) The gametes produced by F_1.
   (*iii*) The genotypes and phenotype of testcross progeny.
6. A tall plant is crossed with a dwarf plant. In the progeny, about one-half of the plants are tall and the remaining one-half are dwarf. Determine the genotypes of the tall and dwarf plants.
7. In pea, yellow cotyledon is dominant over green cotyledon colour. A plant heterozygous for yellow cotyledon colour is crossed with a plant homozygous for green cotyledon colour. Determine the gametes produced by these plants. and the genotypes and the phenotypes of progeny obtained from their cross.
8. A plant heterozygous for yellow cotyledon colour is self-fertilized. Determine the gametes produced by this plant, and the genotypes and the phenotypes of the progeny.
9. In *Drosophila,* vestigial wings (*vg*) are recessive to normal wings. A normal female is mated with a normal male. Some of the progeny flies have vestigial wings. Determine the genotypes of the above male and female flies, the gametes produced by them and the frequency of vestigial flies in their progeny.
10. A female with normal wings is mated with a male with vestigial wings. Some of their progeny have vestigial wings. Determine the genotypes of the above male and female, the gametes and the progeny produced by them.
11. In *Drosophila,* rose eye colour (*rs*) is recessive to dull red. A female with rose eyes is mated with a male with red eyes. All their progeny have red eyes. Give the genotypes of the above male, the female and the progeny produced by them. If their progeny were intermated among themselves, give the gametes and the progeny that would be produced.
12. A rose eye female is mated with a red eye male and produces some rose eye progeny. Give the genotypes of the male and the female, the gametes produced by them and the genotypes and phenotypes of their progeny.
13. Ebony body colour (*e*) is recessive to the normal grey in *Drosophila.* A homozygous grey female is mated with an ebony male. Their progeny are intermated among themselves. Depict these matings using suitable gene symbols, and give the gametes, and the genotypes and the phenotypes produced from the above matings.
14. In mice, black fur colour (*b*) is recessive to the normal grey. A grey female when mated with a grey male produces a black progeny. Determine the genotypes of the above male and the female, the gametes produced by them, and the genotypes and phenotypes of their progeny.

10

Laws of Mendel
II. Independent Assortment

10.1. INTRODUCTION

Studies on monohybrids enabled Mendel to formulate *the basic concept* of *inheritance;* the essentials of this concept are summarised below.

1. *Development of each character is controlled by a* ***gene****.*
2. *Each gene exists in two alternative forms called* ***alleles.*** *Alleles govern the contrasting forms of a character.*
3. *The genes are particulate. Therefore, the two alleles of a gene do not modify each other when they stay together in the same cell.*
4. *Each somatic cell has two copies of a gene, while gametes have only one copy.*
5. *The two alleles of a gene separate and pass into different gametes of the hybrid.*
6. ***Genes*** *are the units of inheritance passed from one generation to the next.*

These properties explain exceptionally well the 3 : 1 ratio obtained in the F_2 of monohybrid crosses. They were used to formulate the law of segregation.

Mendel also crossed pea varieties, which differed from each other with respect to two characters; such crosses are known as ***dihybrid*** crosses. Based on the studies of such crosses, he proposed the law of independent assortment.

10.2. LAW OF INDEPENDENT ASSORTMENT

According to this law, *the segregation of two or more characters in the same hybrid is independent of each other. Thus any allele of one gene is equally likely to combine with any allele of the other gene and pass into the same gamete. Independent segregation of two genes produces four different types of gametes in equal proportion. A random union among these gametes gives rise to 16 possible zygotes. These zygotes yield a* 9 : 3 : 3 : 1 *phenotypic ratio, which is known as the* ***typical dihybrid ratio.***

Let us suppose that in pea a dihybrid is heterozygous for seed shape (*Ww*) as well as cotyledon colour (*Gg*). We know that segregation for seed shape produces, on an average, 3/4 round and 1/4 wrinkled F_2 seeds. Similarly, segregation for cotyledon colour will give 3/4 yellow and 1/4 green seeds in F_2. The segregation of seed shape and cotyledon colour in the dihybrid will be independent of each other (according to the law of independent assortment). Therefore, *the frequency of any combination of seed shape and cotyledon colour in the* F_2 *generation of the dihybrid*

will be the product of the frequencies in F_2 *of that seed* ***shape*** *and that cotyledon colour.*

Independent Assortment

- Second Law of Mendel.
- When two genes segregate at the same time, their segregation is independent of each other.
- *Consequences* :
  1. An allele of the first gene is equally likely to go into a gamete with an allele of the second gene.
  2. F_1 produces gametes in 1 : 1 : 1 : 1 ratio.
  3. Phenotypic, ratio in F_2 : 9 : 3 : 3 : 1.
  4. Genotypic ratio in F_2 : 1 : 2 : 2 : 4 : 1 : 2 : 1 : 2 : 1.

Thus the frequency of round yellow seeds in the dihybrid F_2 will be 3/4 × 3/4 = 9/16. (Here, the first 3/4 is the frequency of round seeds in F_2, and the second 3/4 is the frequency of yellow seeds in F_2. Similarly, the frequency of round green seeds will be, 3/4 × 1/4 = 3/16; that of wrinkled yellow seeds will be, 1/4 × 3/4 = 3/16; and that of wrinkled green seeds will be, 1/4 × 1/4 = 1/16 (Table 10.1).

TABLE 10.1
The Kinds and Frequencies of Different Phenotypes Expected in F_2 from Independent Segregation for Seed Shape and Cotyledon Colour in Pea

Segregation for Ww		*Segregation for* Gg		*Independent segregation for seed shape and cotyledon colour*	
Seed shape	*Frequency (= probability)*	*Cotyledon colour*	*Frequency (= probability)*	*Phenotype*	*Frequency (= probability)*
Round	3/4	Yellow	3/4	Round yellow	$\frac{3}{4}\times\frac{3}{4}=\frac{9}{16}$
Wrinkled	1/4	Green	1/4	Round green	$\frac{3}{4}\times\frac{1}{4}=\frac{3}{16}$
Ratio	3 : 1	Ratio	3 : 1	Wrinkled yellow	$\frac{1}{4}\times\frac{3}{4}=\frac{3}{16}$
				Wrinkled green	$\frac{1}{4}\times\frac{1}{4}=\frac{1}{16}$
				Ratio	9 : 3 : 3 : 1

Thus the expected frequencies of the four combinations of seed shape and cotyledon colour are as follows: round yellow 9/16, round green 3/16, wrinkled yellow 3/16, and wrinkled green 1/16. These frequencies may be summarised as 9 round yellow, 3 round green, 3 wrinkled yellow and 1 wrinkled green, or more simply as 9 : 3 : 3 : 1.

10.3. INHERITANCE OF SEED SHAPE AND COTYLEDON COLOUR IN PEAS

Mendel crossed a variety of pea, which had round seeds and yellow cotyledons, with another that had wrinkled seeds and green cotyledons. The F_1 seeds from this cross had round seeds and yellow cotyledons, since round is dominant over wrinkled and

yellow is dominant over green. In the F_2 generation, the following ratio was obtained: 9.84 round yellow : 3.37 round green : 3.12 wrinkled yellow : 1 wrinkled green. This is a close approximation of the 9 : 3 : 3 : 1 ratio. It is the same ratio expected from independent assortment.

Mendel studied the F_3 seeds produced by each F_2 plant. On an average, one out of the 9 plants derived from round yellow F_2 seeds produced only round yellow seeds in F_3, two produced round yellow and wrinkled yellow seeds in 3 : 1 ratio, two other plants produced round yellow and round green seeds in 3 : 1 ratio, while the remaining four produced all the four classes of seeds in the ratio 9 : 3 : 3 : 1. Out of the plants produced by the three round green F_2 seeds, one yielded only round green seeds in F_3, while the remaining two produced round green and wrinkled green seeds in 3 : 1 ratio. Similarly, the plants obtained from the three wrinkled yellow F_2

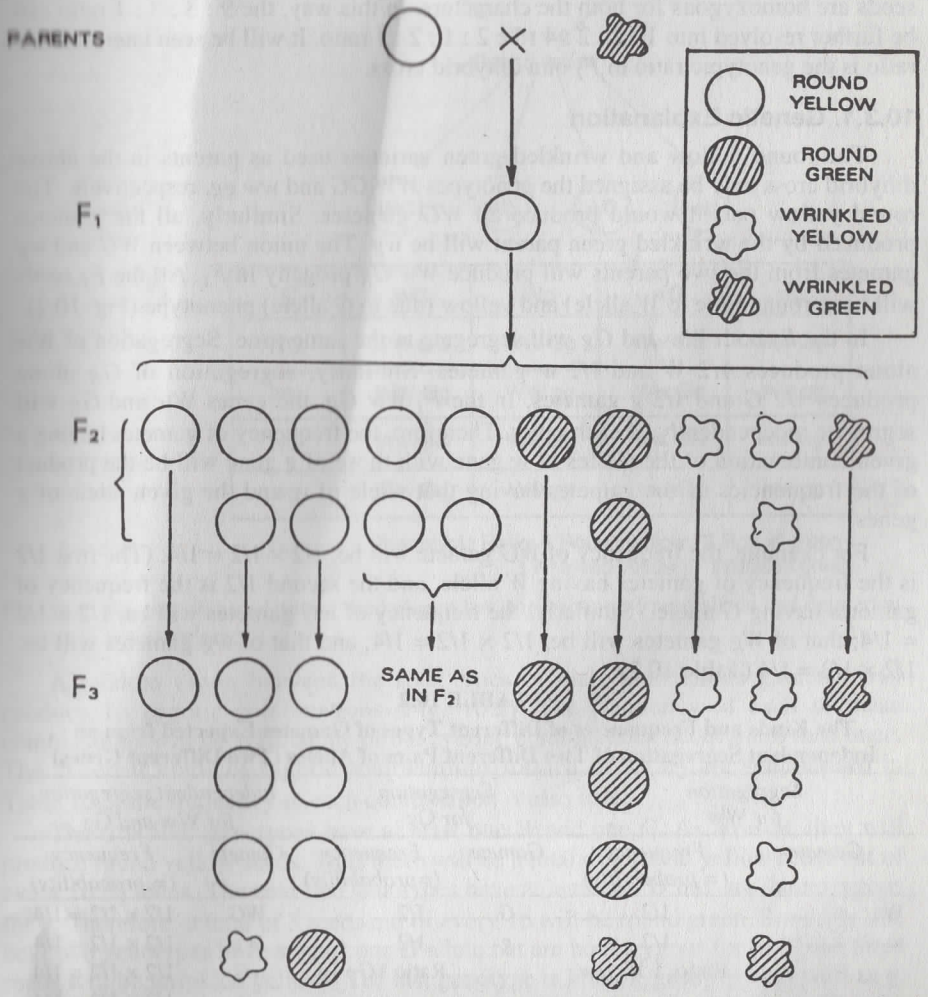


Fig. 10.1. A diagramatic representation of the results obtained by Mendel from a cross between a pea variety having round yellow seeds and a variety with wrinkled green seeds.

seeds were also of two types: one produced only wrinkled yellow seeds in F_3, while two produced wrinkled yellow and wrinkled green seeds in 3 : 1 ratio. Lastly, all the wrinkled green F_2 seeds produced only wrinkled green seeds in F_3 (Fig. 10.1).

On the basis of F_3 seeds produced by them, the 9 round yellow F_2 seeds may be grouped into four categories: 1 seed homozygous for both seed shape and cotyledon colour, 2 seeds homozygous for seed shape but heterozygous for cotyledon colour, 2 seeds heterozygous for seed shape but homozygous for cotyledon colour, and 4 seeds heterozygous for both the characters. The three round green F_2 seeds consist of 1 seed homozygous for both the characters and 2 seeds heterozygous for seed shape but homozygous for cotyledon colour. Similarly, of the three wrinkled yellow seeds 1 seed is homozygous for both the traits and 2 seeds are homozygous for seed shape but heterozygous for cotyledon colour. In the end, all the wrinkled green F_2 seeds are homozygous for both the characters. In this way, the 9 : 3 : 3 : 1 ratio can be further resolved into 1 : 2 : 2 : 4 : 1 : 2 : 1 : 2 : 1 ratio. It will be seen later that this ratio is the genotypic ratio in F_2 of a dihybrid cross.

10.3.1. Genetic Explanation

The round yellow and wrinkled green varieties used as parents in the above dihybrid cross may be assigned the genotypes *WW GG* and *ww gg*, respectively. The round yellow parent would produce all *WG* gametes. Similarly, all the gametes produced by the wrinkled green parent will be *wg*. The union between *WG* and *wg* gametes from the two parents will produce *Ww Gg* progeny in F_1. All the F_1 seeds will have round (due to *W* allele) and yellow (due to *G* allele) phenotype (Fig. 10.2).

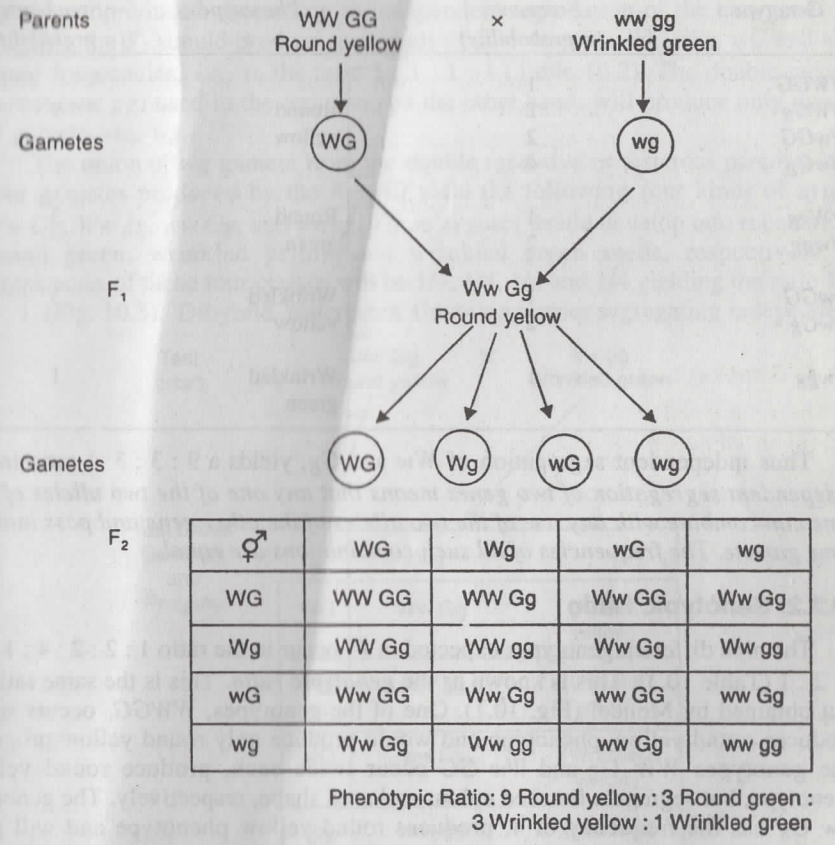
In the F_1 both *Ww* and *Gg* will segregate at the same time. Segregation of *Ww* alone produces 1/2 *W* and 1/2 *w* gametes. Similarly, segregation of *Gg* alone produces 1/2 *G* and 1/2 *g* gametes. In the F_1 *Ww Gg*, the genes *Ww* and *Gg* will segregate independently of each other. Therefore, the frequency of gametes having a given combination of the alleles of *w* gene with those of *g* gene will be the product of the frequencies of the gametes having that allele of *w* and the given allele of *g* genes.

For example, the frequency of *WG* gamete will be, 1/2 × 1/2 = 1/4. (The first 1/2 is the frequency of gametes having *W* allele, and the second 1/2 is the frequency of gametes having *G* allele.) Similarly, the frequency of *wG* gametes will be, 1/2 × 1/2 = 1/4; that of *Wg* gametes will be, 1/2 × 1/2 = 1/4; and that of *wg* gametes will be, 1/2 × 1/2 = 1/4 (Table 10.2).

TABLE 10.2
The Kinds and Frequencies of Different Types of Gametes Expected from an Independent Segregation of Two Different Pairs of Alleles (Two Different Genes)

Segregation for Ww		*Segregation for* Gg		*Independent segregation for* Ww *and* Gg	
Gamete	*Frequency (= probability)*	*Gamete*	*Frequency (=probability)*	*Gamete*	*Frequency (= probability)*
W	1/2	*G*	1/2	*WG*	1/2 × 1/2 = 1/4
w	1/2	*g*	1/2	*Wg*	1/2 × 1/2 = 1/4
	Ratio, 1 *W* : 1*w*		Ratio 1*G* : 1*g*	*wG*	1/2 × 1/2 = 1/4
				wg	1/2 × 1/2 = 1/4
					Ratio, 1 *WG* : 1 *Wg* : 1 *wG* : 1 *wg*

Thus it is clear that during gamete formation, any allele of the gene *w* can combine with any allele of the gene *g*. This gives the following four combinations: *WG*, *Wg*, *wG* and *wg*. The frequencies of these combinations are equal (*i.e.*, 1/4, or 1 : 1 : 1 : 1). This is the law of ***independent assortment.***



♂	WG	Wg	wG	wg
WG	WW GG	WW Gg	Ww GG	Ww Gg
Wg	WW Gg	WW gg	Ww Gg	Ww gg
wG	Ww GG	Ww Gg	ww GG	ww Gg
wg	Ww Gg	Ww gg	ww Gg	ww gg

Phenotypic Ratio: 9 Round yellow : 3 Round green : 3 Wrinkled yellow : 1 Wrinkled green

Fig. 10.2. Results obtained due to the independent segregation of seed shape and cotyledon colour in pea.

A random union between the four types of male and female gametes will produce 16 zygotic combinations (Fig. 10.2). The frequency of each of these combinations will be 1/16. Some of the zygotic combinations occur more than once. The different kinds of zygotic combinations formed in the F_2 are summarised in Table 10.3; the frequency of each combination is also listed.

The first four genotypes have at least one *W* and one *G*. As a result, they will produce round yellow seeds. Thus there will be a total of 9 round yellow seeds out of every 16 F_2 seeds. The next two genotypes have at least one *W* but are homozygous for *g*. Therefore, a total of 3 seeds out of every 16 will be round green. Similarly, the next two genotypes have at least one *G* allele but are homozygous for *w*. These three seeds will be wrinkled yellow. The last genotype is homozygous for *w* as well as *g*. Therefore, the single F_2 seed having this genotype will be wrinkled green. This, in summary, gives the 9 : 3 : 3 : 1 ratio.

TABLE 10.3
The Various Genotypes and Phenotypes and Their Frequencies in F_2 Generation of A Dihybrid Cross for Seed Shape and Cotyledon Colour in Pea; Summarised from Fig. 10.2

Genotype	*Frequency (= probability)*	*Phenotype*	*Frequency (= probability)*
WWGG	1		
WWGg	2	Round yellow	9
WwGG	2		
WwGg	4		
WWgg	1	Round green	3
Wwgg	2		
wwGG	1	Wrinkled yellow	3
wwGg	2		
wwgg	1	Wrinkled green	1

Thus independent segregation of *Ww* and *Gg*, yields a 9 : 3 : 3: 1 ratio in F_2. *Independent segregation of two genes means that any one of the two alleles of one gene can combine with any one of the two alleles of the other gene and pass into the same gamete. The frequencies of all such combinations are equal.*

10.3.2. Genotypic Ratio

The nine different genotypes expected in F_1 occur in the ratio 1 : 2 : 2 : 4 : 1 : 2 : 1 : 2 : 1 (Table 10.3). This is known as the ***genotypic ratio***. This is the same ratio as that obtained by Mendel (Fig. 10.1). One of the genotypes, *WWGG*, occurs once, produces round yellow phenotype and would produce only round yellow progeny. The genotypes *WW Gg* and *Ww GG* occur twice each, produce round yellow phenotype, but segregate for seed colour and seed shape, respectively. The genotype *Ww Gg* has the frequency of 4, produces round yellow phenotype and will give a 9 : 3 : 3 : 1 ratio in F_3. One seed has the genotype *WW gg*, produces round green phenotype and will yield only round green progeny. Two seeds have the genotype *Ww gg* and would segregate for seed shape in F_3. The genotype *ww GG* has the frequency of one, and would produce only wrinkled yellow progeny. The genotype *ww Gg* occurs twice and would show segregation for cotyledon colour in the next generation. The remaining genotype, *ww gg*, occurs only once and would produce all wrinkled green seeds in the next generation. *Thus the results actually obtained by Mendel and those expected from independent segregation of the two concerned genes are in perfect agreement.*

Testcross
- F_1 crossed with double recessive strain.
- *Phenotypic ratio* : 1 : 1 : 1 : 1.
- *Conclusion* : F_1 produces 4 types of gametes in 1 : 1 : 1 : 1 ratio.

10.4. TESTCROSS OF A DIHYBRID

In a ***testcross,*** the F_1 hybrid (*WwGg*) is crossed with a variety or strain having the recessive forms of both the characters (***double recessive***), *i.e.*, wrinkled green. The purpose of a testcross is to verify the ratio in which the different gametes of the F_1 hybrid are produced. According to independent segregation of the two genes (*Ww* and *Gg*), the F_1 should produce four kinds of gametes, *viz.*, *WG*, *Wg*, *wG* and *wg*, in equal frequencies, *i.e.*, in the ratio 1 : 1 : 1 : 1 (Table 10.2). The double recessive parent (*ww gg*) used in the testcross, on the other hand, will produce only one kind of gamete, *viz.*, *wg*.

The union of *wg* gamete from the double recessive or testcross parent with the four gametes produced by the F_1 will yield the following four kinds of zygotes: *Ww Gg*, *Ww gg*, *ww Gg*, and *wwgg*. These zygotes would develop into round yellow, round green, wrinkled yellow and wrinkled green seeds, respectively. The frequencies of these four classes will be 1/4, 1/4, 1/4 and 1/4 yielding the ratio 1 : 1 : 1 : 1. (Fig. 10.3). Dihybrid testcrosses (involving genes segregating independently)

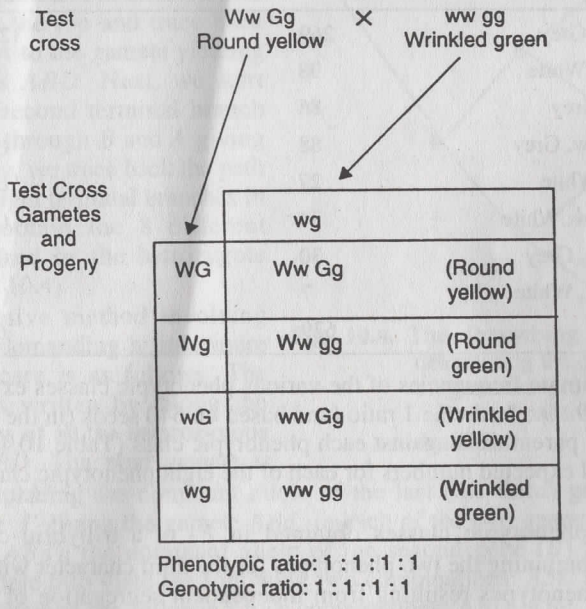


Fig. 10.3. Results from a test cross of an F_1 hybrid heterozygous for seed shape and cotyledon colour (*Ww Gg*).

do actually yield four types of progeny in the ratio 1 : 1 : 1 : 1. In fact, the occurrence of 1 : 1 : 1 : 1 ratio in a dihybrid testcross is accepted as sufficient evidence for independent segregation of the concerned genes.

Trihybrid Ratio
- Three genes segregate in the same F_1 hybrid.
- Their segregation is independent.
- Phenotypic ratio:
  1. F_2 : 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1.
  2. Testcross: 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1.

10.5. TRIHYBRID RATIO

Independent assortment in a trihybrid (hybrid obtained by crossing two parents differing for three characters) produces 8 different kinds of gametes (Fig. 10.4). These male and female gametes will unite randomly to yield a total of 64 zygotic combinations, which can be classified into eight different phenotypic classes. This gives the ratio of 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1. In F_2 of a trihybrid cross segregating for seed shape, cotyledon colour and seed coat colour, Mendel obtained the following data (the total number of seeds was 639).

TABLE 10.4
Frequencies of Different Phenotypic Classes Obtained by Mendel in the F_2 Generation of A Trihybrid Cross (Segregation for Seed Shape, Cotyledon Colour and Seed Coat Colour)

Phenotypic class	*Observed frequency by Mendel*	*Expected Frequency on the basis of 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1 ratio)*
Round, Yellow, Grey	269	(270)
Round, Yellow, White	98	(90)
Round, Green, Grey	86	(90)
Wrinkled, Yellow, Grey	88	(90)
Round, Green, White	27	(30)
Wrinkled, Yellow, White	34	(30)
Wrinkled, Green, Grey	30	(30)
Wrinkled, Green, White	7	(10)
Total	639	640

The approximate frequencies of the various phenotypic classes expected on the basis of 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1 ratio (and based on 640 seeds (in the place of 639) are listed within parenthesis against each phenotypic class (Table 10.4). Obviously, the observed and expected numbers for each of the eight phenotypic classes are quite close.

The eight phenotypic classes obtained in F_2 of a trihybrid can be easily determined by combining the two phenotypes for the third character with each of the four different phenotypes resulting from independent segregation of the first two characters. Similarly, the three genotypes for the third gene may be combined with each of the 9 genotypes obtained from independent segregation of the first two genes to yield the 27 genotypes resulting from a trihybrid cross.

10.6. DETERMINATION OF DIFFERENT GAMETES PRODUCED BY INDEPENDENT SEGREGATION

The different gametes produced due to independent assortment can be determined by a simple 'branching method' (Fig. 10.4). Suppose, an individual is heterozygous for three genes, *Aa, Bb* and *Dd*. A gamete produced by this individual will receive either *A* or *a* allele due to segregation of *Aa*. This is represented by two branches, one for *A* and one for *a*, from a small circle. Any one of the two alleles of *b* gene

(*B* and *b*) may pass into the gamete having *A*; this is shown by two branches, one each for *B* and *b*, arising from *A*. The same is the case for *a* as well since either *B* or *b* allele may be included in this gamete. This also is represented by one branch each for *B* and *b* arising from *a* (Fig. 10.4). Similarly, any one of the two alleles at locus *d* (*D* and *d*) will be transmitted with each of the *B* and *b*. This is also represented by two branches (one for *D* and one for *d*) arising from each *b* and *B* allele.

The genotype of a gamete is arrived at by starting at each of the terminal branches and tracing its path back to the origin point or the circle, which represents the gamete. For example, we start with *D* at the top and trace back through *B* and *A* to the gamete yielding its genotype as *ABD*. Next, we start from *d* at the second terminal branch and trace back through *B* and *A* giving *ABd*. In this way, we trace back the path of each of the eight terminal branches in Fig. 10.4 to obtain the 8 different gametes produced by the hetrozygote *Aa Bb Dd* (Fig. 10.4).

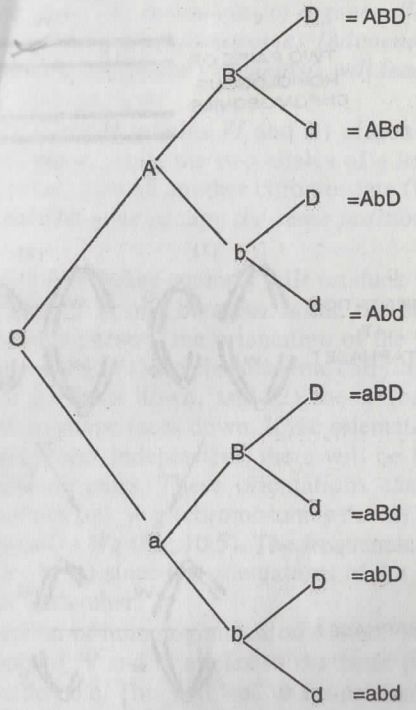


Fig. 10.4. The 'branching method' of determining the different types of gametes produced by an individual with the genotype *Aa Bb Dd*.

An alternative method involving less work but demanding a little more attention and care is as follows. The first gamete is written by taking the dominant alleles of all the heterozygous genes, *e.g., ABD*. The next gamete is obtained by replacing the dominant allele of the last (the third) gene *D* with its recessive allele *d*, giving the gamete *ABd*. In each of the two gametes obtained so far (*ABD* and *ABd*), the dominant allele of the second gene (*B*) is replaced by its recessive allele *b* to yield the following two combinations: *AbD* and *Abd*. Thus far, we have obtained the following four combinations: *ABD, ABd, AbD* and *Abd*. Finally, we replace the dominant allele of the first gene (*A*) with its recessive allele *a* in all the above four combinations. This gives, *aBD, aBd, abD* and *abd*. Thus we get the same eight gametes by this procedure as well as those obtained from the branching method.

Independent Assortment: Physical Basis
- The two genes are located in two different chromosomes.
- The two chromosomes orient independent of each other at *MI*.
- This gives 4 possible orientations in equal frequency.
- This produces the 4 types of gametes in equal frequency.
- This ultimately results in the 9 : 3 : 3 : 1 ratio in F_2.

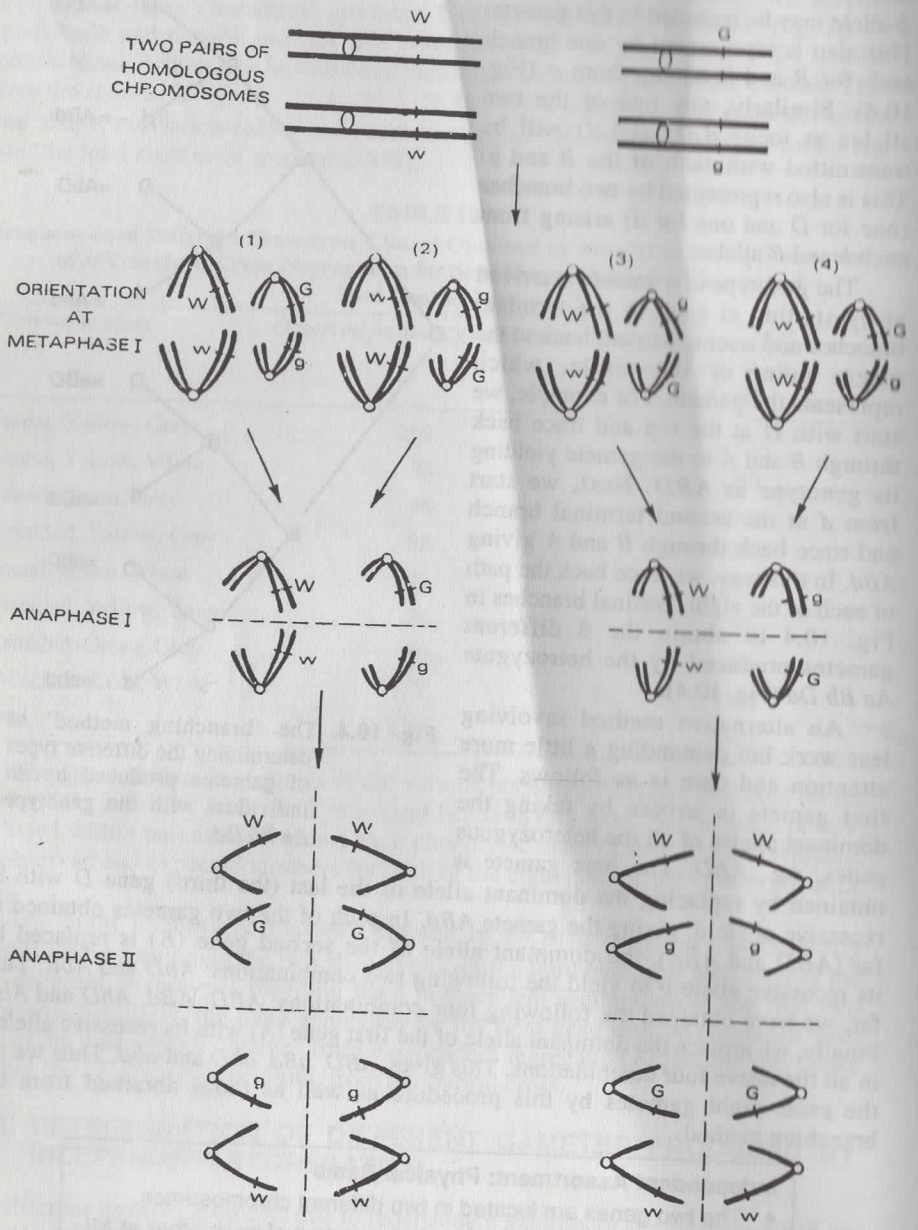


Fig. 10.5. Independent segregation of two genes *Ww* and *Gg* as a consequence of separation of the two pairs for homologous chromosomes, in which *w* and *g* are located, at anaphase I of meiosis.

10.7. PHYSICAL BASIS OF INDEPENDENT ASSORTMENT

Independent segregation for two genes can be explained by assuming that the two genes are located in two different chromosomes. The two alleles of a gene will be located in the two homologues of the concerned chromosome. Independent separation of these two pairs of chromosomes at anaphase I of meiosis will lead to the independent segregation of the genes located in them.

Let us assume that the two alleles of *w* locus (alleles *W* and *w*) of pea are located in the two homologues of a chromosome, while the two alleles of *g* locus (alleles *G* and *g*) are situated in the two homologues of another chromosome (Fig. 10.5). *It may be emphasised that the alleles of a gene occupy the same position in the two homologues of a chromosome.*

Pairing between homologous chromosomes during meiosis will produce two bivalents. The alleles *W* and *w* will be present in one bivalent, while the other bivalent will carry the alleles *G* and *g*. At metaphase I, the orientation of the two chromosomes in a bivalent can be of two types, *e.g.,* (1) the chromosome carrying *W* (or *G*) faces up and that carrying *w* (or *g*) faces down, and (2) the *w* (or *g*) chromosome faces up and the *W* (or *G*) chromosome faces down. If the orientations of the two chromosome pairs at metaphase I are independent, there will be four possible orientations of the two chromosome pairs. These orientations can be summarised as follows: (1) *W G* (chromosomes up): *w g* (chromosomes down), (2) *w g* (up) : *WG* (down), (3) *Wg* : *wG* and (4) *wG* : *Wg* (Fig.10.5). The frequencies of these four orientations will be equal (1 : 1 : 1 : 1) since the orientations of the two chromosome pairs at MI are independent of each other.

In the first two orientations, the separation of homologous chromosomes at AI (anaphase I) will result in the transmission of *W* and *G* alleles to the same pole, while *w* and g alleles will pass to the opposite pole. This will lead to the production of two *WG* and two *wg* gametes from each cell after the completion of meiosis. The separation of homologous chromosomes in the case of the last two orientations will lead to the transmission of *W* and *g* alleles to one pole, while *w* and *G* alleles will move to the opposite pole. The completion of meiosis in each such cell will produce two *Wg* and two *wG* gametes (Fig. 10.5). *Thus an independent separation of two pairs of chromosomes, each pair carrying alleles of one of the two genes, will lead to an independent segregation of the alleles of the two genes.*

QUESTIONS

1. Explain the law of independent assortment with the help of a suitable example. Describe the results obtained from a testcross of a dihybrid F_1.
2. Discuss the physical basis of independent segregation of two genes with the help of suitable diagrams.
3. In snapdragons, normal flowers are dominant over peloric ones and tall plants are dominant over short ones. If a homozygous tall, normal-flowered plant is crossed with a dwarf, peloric-flowered one, give the genotype and phenotype of F_1, the gametes produced by F_1 and the different genotypes and phenotypes obtained in F_2 and their frequencies.
4. In snapdragons, red flower is partially dominant over white, the heterozygotes having pink flowers. A plant homozygous for white, normal flowers is crossed with another plant homozygous for red, peloric flowers. Give the genotype and phenotype of F_1, and the different genotypes in F_2 having the phenotype red, peloric flowers.

5. A dwarf pea plant with yellow seeds is crossed with a tall plant with green seeds. Give the genotype and phenotype of F_1, the gametes produced by F_1, the genotypes and phenotypes of F_2 and testcross progeny.
6. In guinea pigs, rough coat is dominant over smooth coat and black coat is dominant over white coat; the two characters are governed by genes located in two different chromosomes. A homozygous rough, black animal is crossed with a smooth, white one. Give the genotypes and phenotypes of $F1$, the gametes produced by F_1, and the genotypes and phenotypes obtained in F_2 and testcross generations.
7. An individual having three pairs of chromosomes has received the centromeres *A, B,* and *D* from his father and the centromeres *a, b* and *d* from his mother. List the various combinations of the centromeres obtaining in its gametes. (*Note*: consider the centromeres as genes).
8. Determine the different gametes produced by the following:
   (*i*) *Aa Bb,*
   (*ii*) *Aa bb,*
   (*iii*) *AA Bb*
   (*iv*) *Aa Bb Dd,*
   (*v*) *Aa Bb Ee,*
   (*vi*) *AA Bb Dd,*
   (*vii*) *Dd Ee Ff,*
   (*viii*) *Aa Bb dd.*
9. In *Drosophila,* vestigial wings (*vg*) are recessive to normal wings, and sepia eye colour (*se*) is recessive to red eyes. Homozygous normal females are mated with vestigial, sepia males. The resulting progeny are mated among themselves. Depict these crosses using suitable gene symbols, and giving the gametes, and the genotypes and phenotypes of the progeny produced.
10. A normal *Drosophila* male is mated with a vestigial sepia male. All their progeny have normal wings, but some of them have sepia eyes. Determine the genotypes of the above male and female, and give the genotypes and phenotypes of their progeny.
11. A normal female fly is mated with a normal male fly. Some of their progeny have vestigial wings and sepia eyes. Determine the genotypes of the above male and female, their gametes, and the genotypes and phenotypes of their progeny.
12. In *Drosophila,* bent wings (*bt*) are recessive to normal wings and purple eye colour (*pr*) is recessive to red eye colour. A normal female is mated with a normal male. All of their progeny have normal wings, but some of them have purple eyes. Determine the genotypes of the above male and female, the gametes produced by them, and the genotypes and phenotypes of their progeny.
13. A normal male fly is mated with a normal female fly. All their progeny have normal red eyes, but some of them have bent wings. But when the above male is mated with a female having bent wings and purple eyes, some of their progeny have bent wings and purple eyes. Determine the genotypes of the above male and the two females. Depict the two crosses using suitable gene symbols, and give the gametes produced, and the genotypes and phenotypes of their progeny.
14. When a normal male fly is mated with the female fly A, all their progeny have grey body, but some have rose eye colour (*rs*). But when the above male is mated with the normal female B, some of their progeny have black body (*b*) and rose eyes. Determine the genotypes of the above male and females A and B. Depict the two crosses using suitable gene symbols, and give the gemetes produced by the parents and the genotypes and phenotypes of their progeny. (*Note:* black body colour is recessive to grey, and rose eye colour is recessive to red.)

11

Physical Basis of Heredity (Chromosomal Theory of Inheritance)

11.1. INTRODUCTION

In the year 1900, Mendel's paper was rediscovered independently by Correns, de Vries and Tschermak. This paper of Mendel outlined the essential features of the behaviour of genes during inheritance. In the same year, the second edition of '*The Cell in Development and Inheritance*' by Wilson was published. This book brilliantly summarised and elaborated the various aspects of chromosomes discovered till that time. *It was generally supposed that genes were located on chromosomes, but a critical assessment of the supposition was lacking.*

11.2. THE CHROMOSOMAL THEORY OF INHERITANCE

The idea of chromosomal basis of segregation and independent assortment was put forth by Sutton in 1902. This hypothesis was elaborated more fully and was critically evaluated in his 1903 paper. This paper marks the coming together of genes of Mendel and a well established cell organelle, the chromosomes. This synthesis gave us the science of cytogenetics. *The hypothesis that genes are located in chromosomes was designated as* ***Sutton-Boveri hypothesis*** or, more commonly, as the ***chromosomal theory of inheritance***. As indicated above, Boveri is given equal credit for this theory as he had independently reached a similar conclusion in the same year.

11.3. EVIDENCE FOR THE THEORY

That the genes are located in chromosomes is accepted as a matter of fact today. It is no more regarded as a theory or hypothesis. The various direct and indirect evidences that establish the location of genes in chromosomes are summarised below.

11.3.1. Parallels in the Behaviours of Genes and Chromosomes

1. Each somatic cell contains two copies of each gene. Similarly, each somatic cell has two copies or homologues of each chromosome.
2. Each gamete contains only one copy of a gene or only one allele of a gene. The phenomenon that ensures this is termed as segregation. Similarly, each gamete receives only one of the two homologues of each chromosome pair.
3. In Mendel's and several subsequent studies, two genes assorted independently to yield typical dihybrid ratio. Similarly, different chromosome pairs or bivalents orient themselves at metaphase I independent of each other. If two genes are located in two different

chromosomes, random assortment of the two chromosome pairs at anaphase I would adequately explain independent assortment of the two genes (Fig. 10.5).

4. In 1901, McClung discovered the accessory or *X* chromosome of grasshopper and postulated that this chromosome was involved in sex determination. This was the first demonstration that a character (in this case sex) was associated with a specific chromosome.

The evidence presented thus far is quite indirect and relies on parallels in the behaviours of genes and chromosomes.

11.3.2. Sex Linkage

In 1910, Morgan presented a more direct evidence supporting the chromosome theory of heredity. He reported the peculiar inheritance pattern of *white* eye (*w*) gene of *Drosophila.* One of the unusual features of this inheritance was as follows: when white-eyed females were mated with red eyed males, all the F_1 females had red eyes but all the F_1 males were white-eyed (Figs. 20·1, 20·2). The pattern of transmission of *white* eye gene was identical with that of the *X* chromosome of *Drosophila.* This prompted Morgan to postulate that the gene *w* was located in the *X* chromosome. *This was the first demonstration of an association between a specific gene and a specific chromosome.*

11.3.3. Non-disjuncion of *X* Chromosome

In 1913, Bridges observed an unusual transmission of a sex-linked gene in *Drosophila.* When vermilion-eyed females were mated with red-eyed males, a small frequency (1 in 2-3 thousand flies) of female flies in F_1 had vermilion eye colour in place of the normally expected red colour. The unusual transmission of the *vermilion* gene was explained on the basis of nondisjunction of the *X* chromosome in some of the oocytes. Such a nondisjunction gives rise to some definite cytological expectations (Fig. 11.3) The chief cytological expectation is that the unusual F_1 females, *i.e.*, those having vermilion eye colour, should have *XXY* chromosome constitution. This as later confirmed by Bridges through cytological analysis of the unusual flies. *This was a concrete proof in fovour of the chromosomal theory of inheritance.*

11.3.4. Attached *X* Chromosome

In 1922, L.V. Morgan discovered a female fly, which regularly produced the unusual transmission of a sex-linked gene (*yellow* body colour). She concluded that the two *X* chromosomes of this fly were attached with each other. Cytological analysis of the progeny confirmed this expectation. This further supported the conclusion that the sex-linked genes of *Drosophila* are located in the *X* chromosome.

11.3.5. Linkage and Chromosome Maps

1. In 1910, Morgan described the phenomenon of linkage and recombination (crossing over) between two sex-linked genes in *Drosophila.* A little later, he described linkage between some more sex-linked genes of *Drosophila* and proposed that the strength of linkage between any two genes depends on the distance between them in the chromosome. In other words, the frequency of recombination between two genes depends on the distance between them in the chromosome. This idea was soon extended by Sturtevant who proposed the preparation of chromosome maps based on the frequencies of recombination between linked genes. He prepared the first chromosome map of some sex-linked genes of *Drosophila.* This technique was soon exploited to prepare chromosome maps of several organisms.
2. Extensive linkage maps have been prepared in several organisms, *e.g.*, *Drosophila,* maize, barley, *Neurospora,* yeast, peas, etc. In each such case, *the number of different linkage groups is the same as the number of chromosomes in the haploid set.* Not a single case in known where the two differ. *This is exactly what would be expected if genes were located in chromosomes.*
3. Morgan explained the recombination between linked genes to be due to the exchange of chromatin material between homologous chromosomes. In 1931, Stern (in *Drosophila*) and McClintock (in maize) demonstrated that *recombination between linked genes indeed involved the exchange of chromatin material between homologous chromosomes. This was a rather direct demonstration that genes are located in chromosomes.*
4. If genes are located in chromosomes, structural changes in chromosomes should lead to a corresponding change in the linkage relationships of the concerned genes. In practice, this is always the case. For example, when a segment of a chromosome is inverted, the inversion homozygote shows an inverted gene sequence corresponding to the inverted segment. Similarly, in reciprocal translocation homozygotes a corresponding change in the two linkage groups is also obtained. In fact, translocation is used as a handy tool in assigning linkage groups to different arms of individual chromosomes. *Since structural changes in chromosomes are always accompanied by corresponding changes in linkage groups, genes must be located in chromosomes.*
5. *Linkage maps of bacteria and viruses were demonstrated to be circular.* This led to the expectation that both bacterial and viral chromosomes would also be circular. The isolation of intact bacterial and viral chromosomes has confirmed this expectation.

11.3.6. The *Bar* Eye Locus of *Drosophila*

In all *Drosophila flies* having *bar* eyes, the *X* chromosome (at least one *X* chromosome in females) has two 16*A* regions lying side-by-side (*tandem repeat*). This correlation is a perfect one, that is, it is found in all the individuals having *bar* eyes. The presence of three 16*A* segments in a single *X* chromosome produces the ultrabar phenotype. Thus *a genetic effect was localised to a distinctly recognisable small segment of a chromosome.*

11.3.7. Numerical Changes in Chromosomes

1. In addition to structural changes in chromosomes, numerical chromosome changes also furnish evidence in favour of the chromosome theory of heredity. In most cases, a loss (*monosomy*) or a gain (*trisomy*) of a single chromosome produces marked morphological and physiological abnormalities. In a monosomic for a specific chromosome, genes located in that chromosome should be present in a single copy. As a result, in

monosomics such genes show a marked deviation from the normal inheritance pattern. The same is the case for trisomic plants since they carry three copies of the concerned genes in place of the normal two. This property of monosomics and trisomics is exploited in assigning the different linkage groups to individual chromosomes of the haploid set.

2. In a large number of diploid plants, autotetraploids, autotriploids etc. have been experimentally induced. Some plants are known to be natural autoployploids, *e.g.*, alfalfa, potato, coffee, sweet potato, banana, etc. All such plants would be expected to contain more than two copies of all the genes: three copies in triploids, four in tetraploids and so on. Consequently, they are expected to show a more complex inheritance pattern than diploids for even monogenic traits. In fact, natural tetraploids show a complex pattern of inheritance, which can be explained only on the basis of the presence of four copies of the concerned genes.

11.3.8. Y-Linked Inheritance

Some genes, *e.g.*, *SRY* gene in humans, are located in the *Y* chromosome. Such genes are transmitted from father to son and are absent in females. Characters produced by *Y*-linked genes are known as ***holandric characters*** since they are confined to males only.

11.3.9. Genetic Material

It is accepted beyond any reasonable doubt that DNA is the genetic material in eukaryotes and in prokaryotes, except in some viruses. It, therefore, follows that the organelles containing DNA must contain genes and *vice-vèrsa*. By far, the vast majority of DNA occurs in chromosomes of eukaryotes. It is, therefore, self-evident that genes must be an integral part of eukaryotic chromosomes.

11.3.10. *In Situ* Hybridization

Gene expression occurs as follows: (1) RNA polymerase produces RNA copies (mRNA) of one strand of the gene, (2) the mRNA directs synthesis of a specific protein, and (3) this protein participates in cellular activities to produce the phenotype specified by the gene. Radioactively labelled DNA copies of different mRNA molecules have been shown to base-pair (***hybridize***) with DNA in different but specific locations of chromosomes; this technique is called ***in situ hybridization***. This demonstrates beyond any doubt that genes are present in chromosomes.

11.3.11. Transgenic Organisms

The mRNA produced by a gene can be used to obtain copies of the concerned gene as follows: (1) the enzyme reverse transcriptase can be used to produce a DNA copy (***cDNA***) of the mRNA using the latter as a template. (2) Alternatively, the genomic DNA (=DNA present in chromosomes) of the organism may be isolated and the genomic DNA segment complementary to the mRNA may be obtained. Genes obtained by either technique have been transferred into a different organism, where they are expressed. Copies of a gene obtained by both the approaches produce identical proteins. This clearly shows that the genes are located in chromosomes.

The foregoing considerations clearly demonstrate that there is little doubt that genes are located in chromosomes. This conclusion is universally accepted as one of the fundamental facts of genetics and cytogenetics.

QUESTIONS

1. Briefly describe the various evidences, which prove that genes are located in chromosomes.
2. Write short notes on the following: (*i*) chromosome theory of heredity, (*ii*) nondisjunction of *X* chromosome and chrmosome theory of heredity, (*iii*) attached-X and the chromosome theory of inheritance, (*iv*) structural chromosome aberrations and chromosome theory of heredity, and (*v*) genetic material and chromosome theory of heredity.

12

Gene Symbols, Dominance, and Lethal Genes

12.1. GENE SYMBOLS

For a clear and brief discussion on the inheritance of various characters and of other genetic concepts, it is necessary to represent the genes governing various characters by simple, brief and distinct symbols. This need was felt by Mendel himself, who used *Aa, Bb* etc. to designate different traits of pea. For example, he designated the round seed shape with *A*, wrinkled seed shape with *a*, yellow cotyledon colour with *B*, and green cotyledon colour with *b*.

The use of gene symbols is necessary because (*i*) it saves time, labour and space, (*ii*) it allows a clear description of the inheritance of those traits, which are governed by more than one gene, (*iii*) the concept of genotype can be properly elaborated and (*iv*) brief symbols permit the preparation of brief, clear and comprehendible genetic maps etc. (these would be immensely cumbersome if character names were to be written in place of gene symbols). In fact, gene symbols are so commonplace in genetic literature today that they are taken for granted by every one but the utter novice in the subject.

Gene Symbols

- Derived from the name of the trait governed :
  1. From dominant form of the trait (used initially).
  2. From recessive form of the trait (used subsequently).
  3. From mutant form of the trait (used currently).
- First letter or first letter plus one or two more letters.
- First letter of dominant allele is in capital, while that of recessive allele is in small letters.
- The wild type allele denoted by using + as superscript to the symbol for mutant allele.
- In case of mutiple alleles, letters may be used as superscripts (or numbers may be used as subscripts) of the first discovered mutant allele.

12.2. DERIVATION OF GENE SYMBOLS

The two universal conventions of gene symbols are as follows. (1) The two alternative forms of a single character, which show a monohybrid ratio (3 : 1 or a modification of 3 : 1) on segregation are assigned the two forms of the same letter or

letters, and (2) the capital form of the symbol is assigned to the dominant form of the trait, while the recessive form is assigned the small form of the chosen symbol.

It is customary to derive gene symbols from the names of characters they govern. Gene symbols may be derived from the name of the (*i*) dominant phenotype, (*ii*) recessive phenotype or (*iii*) mutant phenotype.

12.2.1. From the Dominant Phenotype

In the early days of genetics, *the names of dominant forms of characters were used for deriving gene symbols.* For example, the gene producing colour (pigment) in flowers and that producing melanin (colour) in the skin cells of animals was given the symbol *C* (first letter of the word '*colour*'). Similarly, the gene producing red flower colour in snapdragons was given the symbol *R* (*red*). These symbols are derived from the names of the dominant forms of the characters governed by the concerned genes, and many such symbols are in use even today (Table 12.1).

TABLE 12·1

Some Gene Symbols Derived from the Names of the Dominant Forms of the Characters

Character	*Gene symbol*	
	Dominant	*Recessive*
Melanin production (animals)	*C* (***colour***)	*c* (albino)
Coat colour (mice)	*Y* (***yellow***)	*y* (grey)
Colour production (plants)	*C* (***colour***)	*c* (white)
Flower colour	*R* (***red***)	*r* (white)
Seed coat colour (barley)	*B* (***black***)	*b* (white)
Seed coat colour (barley)	*Y* (***yellow***)	*y* (white)

12.2.2. From the Recessive Phenotype

Subsequently, it was felt that for certain reasons it was necessary to use *the names of recessive forms of characters for deriving gene symbols.* As before, the gene symbols consist of the first letter or the first letter and one (*e.g., pr, ra, bw, rb*), or two (*e.g., ras, car*) additional letters from the names of respective recessive phenotypes governed by the concerned genes. The need for using one or two letters in addition to the first letter of the character names is obvious since two or more trait names may begin with the same letter, *e.g.*, raspberry, rose, and ruby eye colours of *Drosophila.*

12.2.3. From the Mutant Phenotype

It is necessary to introduce the concept of wild type at this point. Many organisms most commonly used in genetic studies primarily exist in natural populations, *i.e.*, in the wild state. Genetic studies in these organisms have been conducted on a very small fraction of their wild populations brought under human management. Genetic investigations of these organisms became possible due to the discovery of variants possessing different forms of one or the other trait as compared to that of the original population. Such new variants are known as ***mutants*** and the character form exhibiting deviation from the form found in the original (wild) population is the ***mutant phenotype***. The form of a character found in the original (natural or wild) population is the ***wild phenotype*** or simply the ***wild type***.

It is now, agreed upon to use ***the names of mutant phenotypes for obtaining gene symbols for different traits irrespective of the dominance or recessiveness of the mutant phenotype***. According to this system, the first letter, or the first letter along with one or two other letters, from the name of the mutant phenotype is used as a symbol for the gene controlling the trait. If the mutant phenotype is recessive, all the letters in the symbol for the allele determining the mutant phenotype are small. The same symbol is assigned to the allele governing the normal or wild phenotype but the first letter of this symbol is in the capital form to indicate the dominance of this allele. However, the reverse is the case when the mutant phenotype is dominant to the wild type. Thus the alleles governing *notch* wings in *Drosophila* is given the symbol *N,* while the wild type allele is assigned the symbol *n* (Table 12.2).

As a rule, genes are now named after the mutant phenotypes they produce, and the symbols of the mutant alleles are used to designate the normal alleles of the various genes.

TABLE 12·2

Gene Symbols Derived from Names of the Mutant Forms of Characaters (in *Drosophila*)

Character	*Gene symbol**	
	Mutant	*Wild type*
Eye colour	*w* (***white***)	*W* (red)
Eye colour	*v* (***vermilion***)	*V* (red)
Eye colour	*car* (***carnation***)	*Car* (red)
Eye colour	*rb* (***ruby***)	*Rb* (red)
Body colour	*b* (***black***)	*B* (grey)
Body colour	*y* (***yellow***)	*Y* (grey)
Wing shape	*N* (***notched***)	*n* (normal)

*The first letter in the symbyl of the dominant allele is capital.

12.2·4. From the Features of Encoded Protein

Symbols for many genes have been derived from the mass or some other feature of the protein encoded by the gene. For example, M phase kinase hasd two subunits called p^{34} and p^{45} since the mass of these subunits is 34 and 45 kDa, respectively (Chapter 6, Section 6.7). The genes encoding these subunits are dinoted as p^{34} and p^{45}, respectively. Subunit p^{45} is generally called cyclin due to the cyclical change in its concentration during the cell cycle. Yeast genes encoding cyclins are given the symbol *CLN* (from *cyclin*). Similarly, a gene of becterium *Bacillus thuringiensic* encodes a protein that forms crystals in the spores; this gene is given the symbol *cry* (from *crystal*). There are many other similar examples.

12.2.5. Superscript for the Wild Type Allele

A difficulty in the use of mutant phenotype for deriving gene symbols is that the symbols do not indicate the identities of wild type and mutant alleles. But a slight modification facilitates a quick identification of the mutant allele of an unfamiliar gene. In this system, *genes are assigned symbols derived from the respective mutant phenotypes. The dominance of mutant phenotype is indicated by the first letter of the symbol being in the capital form. The allele producing the wild type phenotype is assigned the same symbol as that for the allele governing the mutant phenotype, but a + (plus) sign is added as its superscript.* For example, the black body and

TABLE 12·3

The Gene Symbol for Wild Type of A Charcter Is Derived by Placing A + as the Superscript of the Symbol for the Mutant Form of the Character (in *Drosophila*).

Character	*Gene symbol**	
	Mutant	*Wild type*
Eye colour	*w* (***white***)	w^+ (red)
	v (***vermilion***)	v^+ (red)
	car (***carnation***)	car^+ (red)
	rb (***ruby***)	rb^+ (red)
Body colour	*b* (***black***)	b^+ (grey)
	y (***yellow***)	y^+ (grey)
Wing shape	*n* (***notched***)	n^+ (normal)

* A gene symbol beginning with a capital letter represents a dominant mutant allele, *e.g.,* for notched wings.

notch wing mutants of *Drosophila* are denoted by the symbols *b* (mutant phenotype recessive) and *N* (mutant phenotype dominant), respectively. The symbols for the wild type alleles of *b* and *N* would be b^+ and N^+, respectively, producing grey body and normal wings, respectively (Table 12.3).

The presence of + sign as a superscript of an allele (*e.g.,* b^+) *denotes it to be the wild type allele, while the absence of this sign* (*e.g.*, b) *makes it the mutant allele. If the gene symbol has a capital letter, e.g.,* N^+/N *(normal and notch wings, respectively), the mutant allele is dominant to the wild type allele. On the other hand, if the gene symbol has only small letters, e.g.,* b^+/b, *the mutant allele is recessive to the wild type allele.* This notation system is the most commonly used system today, particularly in the case of microorganisms, *Drosophila* etc.

12.2.6. Wild Type Allele Designated by a + Sign

If the mutant and the normal alleles of a gene are written together, the wild type allele may be represented simply by a + sign (in place of the + superscript to the symbol for the mutant allele). For example, $b^+ b$, $N^+ N$, $car^+ car$, etc. genotypes may be represented more simply as + *b*, + *N*, + *car*, etc. However *the + symbol by itself does not describe the gene it represents; it specifies the concerned gene only if the mutant allele is also written alongwith this sign.*

12.2.7. Symbols for Multiple Alleles

In such situations, symbols for the wild type and the first discoverd mutant allele are derived as described earlier. The symbols for mutant alleles discoverd subsequently are obtained by adding the first letter or the first letter and one or two more letters from the respective mutant phenotypes as superscripts to the symbols for the respective mutant alleles. For example, coloured fur in rabbits is the wild type, and albino or white fur is the first discovered mutant type; these phenotypes are assigned the symbols *C* and *c*, respectively (on the basis of the dominant phenotype, 'coloured'). Subsequently, two more mutant phenotypes governed by alleles of this gene were discovered and named *himalayan* and *chinchilla.* The alleles determining these phenotypes were given the symbols c^h and c^{ch}, respectively; these symbols were derived by adding the letters *h* and *ch* from himalayan and chinchilla, respectively, as superscripts to the symbol for mutat allele *c* (Table 12.4).

TABLE 12·4
Gene Symbols for Multiple Alleles (for Colour in Rabsits)

Phenotype	*Gene symbol* *
Coloured (full; wild = type)	C
Albino (first discovered mutant phenotype)	c
Himalayan (mutant)	c^h
Chinchilla (mutant)	c^{ch}

*$C > c^{ch} > c^h > c$; > indicates dominance over the allele that follows this symbol.

A more commonly used system for designating multiple alleles is as follows. The symbol for the concerned gene is derived as usual, *e.g.*, symbol *S* for self-incompatibility. The different alleles of the multiple allelic series are denoted by serial numbers placed, usually, as subscripts of this symbol, *e.g.*, S_1, S_2, S_3, S_4, etc. Similarly, cyclin encoding genes of yeast are symbolized as *CLN1*, *CLN2*, etc.

Dominance Relations

- Depends on whether one or hoth the alleles of a gene are expressed in heterozygotes:
  1. *Complete Dominance* : only one allele expressed; F_2 ratio, 3 : 1.
  2. *Incomplete Dominance* : Hetrozygote phenotype intermediate; F_2 ratio, 1 : 2 : 1.
  3. *Codominance* : Both alleles expressed; F_2 ratio, 1 : 2 : 1.
  4. *Overdominance* : Heterozygote phenotype exceeds those of homozygotes; ratio, 1 : 2 : 1.
- Dominant alleles generaly encode functional proteins.
- Recessive alleles often encode nonfunctional proteins.
- The genetic background tends to make a functional allele dominant.

12.3. DOMINANCE RELATIONSHIPS

Mendel found that in heterozygous state only one of the two alleles of a gene was able to express itself, *i.e.*, produce the character. This allele is referred to as the ***dominant allele***. The ***recessive allele***, on the other hand, is unable to express itself in the heterozygous state. Thus the phenomenon of dominance describes the relationship between two alleles of the same gene.

The examples of dominance described by Mendel pertained to complete dominance. Later, it was discovered that in many cases the F_1 is not identical with one of the parents; it may either be intermediate between or different from the two parents. All these situations have been grouped into the followiing four categories: (1) complete dominance, (2) incomplete or partial dominance, (3) codominance and (4) overdominance.

12.3.1. Complete Dominance

In case of ***complete dominance,*** the phenotype produced by heterozygotes is identical with that produced by homozygotes for the concerned dominant allele. The dominant allele in such a situation is known as ***completely*** or ***fully dominant.*** An example of complete dominance occurs for seed shape in peas. Round seed shape is

produced by the dominant allele *W*, while wrinkled shape is determined by its recessive allele *w*. Seeds having the genotype *Ww* are round and indistinguishable from those having the genotype *WW*. As a result, characters showing complete dominance yield the typical 3 : 1 monohybrid ratio in F_2 (Fig. 9.2).

12.3.2. Incomplete Dominance

In many cases, the intensity of phenotype produced by heterozygotes is less than that produced by homozygotes for the concerned dominant allele. Therefore, the phenotype of heterozygotes falls between those of the homozygotes for the two concerned alleles. Such a situation is known as ***incomplete*** or ***partial dominance***, and the dominant allele is termed as ***incompletely*** or ***partially dominant.*** A n example of partial dominance is known for flower colour in four-O'clock (*Mirabilis jalapa*). A partially dominant allele *R* produces red flowers in the homozygous state, while its recessive allele *r* produces white flowers. When a red flowered (*RR*) plant is crossed with one having white flowers (*rr*), the F_1 (*Rr*) has pink flowers. Thus the intensity of flower colour in F_1 (*Rr*) is intermediate between those produced by the two homozygotes (*RR* and *rr*). Segregation in F_2, therefore, produces the phenotypic ratio of 1 red, *RR* : 2 pink, *Rr* : 1 white, *rr*, in the place of 3 : 1, the typical monohybrid ratio (Fig. 12.1).

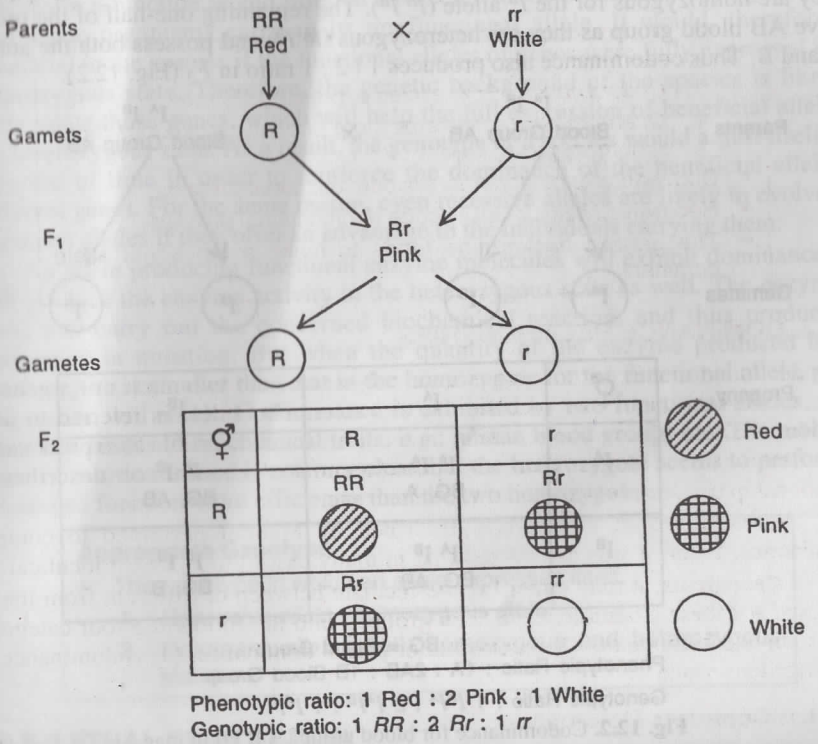


Fig. 12.1. Incomplete domiance for flower colour in Four O'clock (*Mirabilis jalapa*).

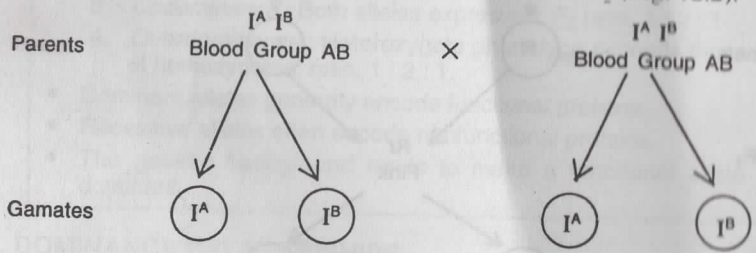
Clearly, heterozygotes for a partially dominant allele have a phenotype, which is easily distinguishable from that produced by the homozygotes for the two alleles. *Therefore, genotypes of plants for characters exhibiting partial dominance can be*

easily deduced from their penotypes. This is not possible in the case of traits exhibiting complete dominance since in such cases heterozygotes have the same phenotype as that exhibited by the homozygotes for the dominant alleles.

12.3.3. Codominance

In ***codominance,*** both the alleles of a gene express themselves in the heterozygotes. As a result, heterozygotes for such genes possess the phenotypes produced by both the concerned alleles. Blood group antigens of man present excellent examples of codominance. One of the most widely known, and the earliest recognised, human blood groups is the ABO blood group. These blood groups arise due to the presence of an antigen on the surface of red blood cells. These antigens are produced by a gene *i*. One dominant allele of this gene, *viz.*, the I^A allele, produces antigen A, which gives rise to blood group A. Another dominant allele of the gene *i, viz.*, I^B, produces antigen B, which is responsible for blood group B. In the heterozygote $I^A I^B$, both the allele I^A and I^B produce their respective antigens. As a result, the heterozygotes have the AB blood group.

Marriages between heterozygotes $I^A I^B$, having blood group AB, produce three types of progeny. One-fourth of their progeny are homozygous for the allele I^A ($I^A I^A$) and have A blood group, while another one-fourth have B blood group since they are homozygous for the I^B allele ($I^B I^B$). The remaining one-half of the progeny have AB blood group as they are heterozygous ($I^A I^B$) and possess both the antigens A and B. Thus codominance also produces 1 : 2 : 1 ratio in F_2 (Fig. 12.2).



Progeny

♂ / ♀	I^A	I^B
I^A	$I^A I^A$ BG, A	$I^A I^B$ BG, AB
I^B	$I^A I^B$ BG, AB	$I^B I^B$ BG, B

BG = Blood Group

Phenotypic Ratio : 1A : 2AB : 1B Blood Group

Genotypic Ratio : 1 $I^A I^A$: 2 $I^A I^B$: 1 $I^B I^B$

Fig. 12.2. Codominance for blood group (A B O) in man.

12.3.4. Overdominance

In case of some genes, the intensity of character expression is greater in the heterozygotes than in the two concerned homozygotes. This situation is known as ***overdminance.*** True overdominance is known in the case of some genes. Many

cases of presumed overdominance may actually be cases of repulsion phase linkages and/or epistasis.

The *white* eye (*w*) gene of *Drosophila* exhibits overdominance for some of the eye pigments, *e.g.*, sepiapteridine and himmelblaus. Allele *w* produces white eye in the homozygous state (*ww*), while its completely dominant allele *W* gives rise to the normal dull red eye colour. Eye pigments sepiapteridine and himmelblaus are present in low concerntrations in *ww* homozygotes, while *WW* homozygotes have relatively higher concentrations of these pigments. However, flies heterozygous for this gene (*Ww*) have an appreciably higher concentrations of these two pigments than the two homozygotes (*WW* and *ww*). Genes showing overdominance would also give rise to a 1 : 2 : 1 ratio in F_2.

12.4. MOLECULAR BASIS OF DOMINANCE

In general, the dominant allele of a gene has the ability to perform a function, while the recessive allele lacks this ability. Most of the genes generate their effects by producing specific enzyme or nonenzyme proteins. Thus the dominant allele of a gene produces a functional protein, while that produced by the recessive allele is either partially functional or nonfunctional. In most cases, the allele with the ability to perform a function would offer an advantage to the organism carrying this allele over those organisms carrying its nonfunctional allele. It would, therefore, be beneficial to the species if the functional alleles of genes were fully expressed in the heterozygous state. Therefore, the genetic background of the species is likely to accumulate those genes, which will help the full expression of beneficial alleles in the heterozygous state. As a result, the genotype of a species would adjust itself over a period of time in order to reinforce the dominance of the beneficial alleles of different genes. For the same reason, even recessive alleles are likely to evolve into dominant alleles if they offer an advantage to the individuals carrying them.

An allele producing functional enzyme molecules will exhibit dominance as it will produce the enzyme activity in the heterozygous state as well. The enzyme, in turn, will carry out the concerned biochemical reaction, and thus produce the phenotype in question. But when the quantity of the enzyme produced by the heterozygote is smaller than that in the homozygote for the functional allele, partial dominance will result. Codominance is exhibited by two functional alleles, and it generally relates to biochemical traits, *e.g.*, human blood groups etc. The molecular basis of overdominance is not very clear, but the heterozygote seems to perform the concerned function more efficiently than the two homozygotes.

Appropriate Genotype

- The genotype in which an allele expresses itself.
  1. *Recessive allele* : homozygous state.
  2. *Dominant allele* : both homozygous and heterozygous states.

12.5. LETHAL GENE ACTION

Some genes affect the survival of those zygotes or individuals in which they are present in the appropriate genotype. ***Appropriate genotype*** will be homozygous state for a recessive allele, and both homozygous and heterozygous states for a dominant allele. The effects of various genes on survival range from an increase in survival,

through no effect to death of all the individuals carrying the gene. All the genes may be grouped into the following five categories on the basis of their effects on survival: (1) lethal, (2) semilethal, (3) subvital, (4) vital and (5) supervital.

Lethal Gene Action

- Such genes affect survival when present in *appropriate genotype.*
- Types:
  1. *Lethal*: All individuals die before they reproduce; usually recessive.
  2. *Semilethal*: Usually less than 10% individuals survive.
  3. *Subvital*: Some individual die.
  4. *Vital*: No effect on survival.
  5. *Supervital*: Increased survival.

12.5.1. Lethal Gene

A ***lethal gene*** causes death of all the individuals carrying this gene in the appropriate genotype before they reach adulthood. *The* ***appropriate genotype*** for an allele producing a recessive effect on survival would be the homozygous state, while for an allele having a dominant effect on survival both homozygous and heterozygous states would be the appropriate genotype.

Most of the lethal genes are ***recessive lethals***. Hence their lethal effect is expressed only when they are in the homozygous state, and the survival of heterozygotes is unaffected. Many recessive lethals affect some other character in addition to their effect on survival; generally their phenotypic effects are dominant over those of their normal alleles.

A recessive lethal affecting coat colour in mice was discovered by the French geneticist Cuenot in 1905. He found that yellow coat colour in mice is produced by a dominant gene *Y*, while its recessive allele *y* determines the normal gray coat colour. Further, all the mice with yellow coat colour were heterozygotes (*Yy*) and he was unable to find a mouse homozygous for the *Y* allele (*YY*). *Later, it was suggested by Castle and Little that the dominant allele* Y *is a recessive lethal, and that it causes the death of homozygous* YY *embryos at an early stage of development.* Subsequently, it was shown that approximately 25% of the embryos of yellow females mated to yellow males are inviable and they failed to develop beyond a very early stage of embryonic development. This confirmed the suggestion of Castle and Little that the dominant *Y* allele is a recessive lethal.

Mating of yellow females with yellow males produces the following zygotes: one-fourth *YY*, (2) one-half *Yy*, and (3) one-fourth *yy*. Homozygous *YY* embryos die at an early stage in the uterus itself. Heterozygous *Yy* embryos develop normally and give rise to yellow mice, while *yy* embryos develop into gray individuals. Thus the phenotypic ratio obtained in the yellow × yellow matings is 2 yellow : 1 gray in place of the typical monohybrid ratio of 3 : 1 (Fig. 12.3). *This ratio* (2 : 1) *is the typical ratio obtained due to the segregation of a recessive lethal gene, which also produces a dominant phenotypic effect.*

Obviously, the *Y* gene in mice has dominant phenotypic effect on coat colour, but is a recessive lethal. The two characterstic features of a recessive lethal are as follows: (1) *they are always present in the heterozygous state since their*

homozygotes do not survive, and (2) *a cross between heterozygotes for a recessive lethal yields a* 2 : 1 *ratio* (instead of the typical 3 : 1 ratio).

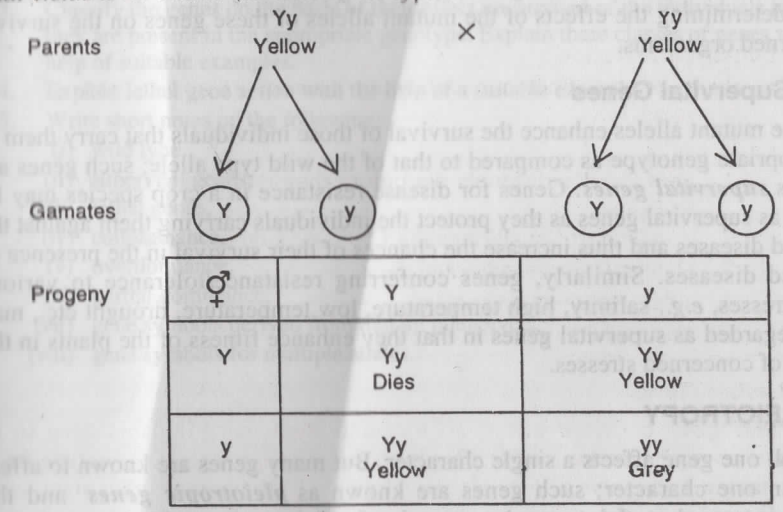


Fig. 12.3. Inheritance of yellow coat colour in mice. Mice with yellow coat colour are always heterozygous.

Many other genes are known that have a dominant phenotypic effect but are recessive lethals, *e.g., aurea* gene (producing yellow leaved plants in the heterozygous state) *in Antirrhinum majus, Dexter* gene in cattle, and *Achondroplastic dwarfness* in man (both producing dwarf condition in the heterozygotes) etc. But many genes are recessive both in their phenotypic as well as lethal effects, *e.g.,* the genes producing albina seedlings in plants, such as, barley.

12.5.2. Semilethal Gene

Semilethal genes do not lead to the death of all the individuals that carry them in the appropriate genotype. However, they cause the death of more than 90% of the individuals; only less than 10% of the individuals survive. Certain *xantha* mutants of many plants are semilethal in the homozygous state.

12.5.3. Subvital Gene

Many mutant genes reduce viability of the individuals carrying them in the appropriate genotype as compared to that of the normal individuals. Such genes kill much less than 90% of the individuals in which they are present. Most of the mutant genes are subvital in their effect, *e.g., miniature* wings in *Drosophila,* some *viridis* mutants of barley etc.

12.5.4. Vital Gene

Those genes, which do not affect the survival of individuals in which they are present in appropriate genotypes, are known as ***vital*** genes. Such genes neither enhance nor reduce the viability of the individuals carrying them. It does not imply that these genes are not necessary for survival of the concerned organisms. It simply means that the survival of the organism is not affected by the fact that the concerned

vital genes are present in homozygous or heterozygous state. Wild type alleles of all the genes of an organism are regarded as vital genes; they serve as the reference point in determining the effects of the mutant alleles of these genes on the survival of concerned organisms.

12.5.5. Supervital Genes

Some mutant alleles enhance the survival of those individuals that carry them in the appropriate genotype as compared to that of the wild type allele; such genes are known as ***supervital genes***. Genes for disease resistance in a crop species may be regarded as supervital genes as they protect the individuals carrying them against the concerned diseases and thus increase the chances of their survival in the presence of concerned diseases. Similarly, genes conferring resistance/tolerance to various abiotic stresses, *e.g.*, salinity, high temperature, low temperature, drought etc., may also be regarded as supervital genes in that they enhance fitness of the plants in the presence of concerned stresses.

12.6. PLEIOTROPY

In general, one gene affects a single character. But many genes are known to affect more than one character; such genes are known as ***pleiotropic genes*** and the condition is termed as ***pleiotropy***. An example of a pleiotropic gene in human beings is the recessive gene *s* (sometimes denoted as Hb^s), which produces sickle cell anemia in the *ss* homozygotes. More than 50% of the individuals homozygous for this gene (*ss*) die before the age of 20 years. The primary effect of this gene is the substitution of a valine molecule for a glutamic acid molecule at position 6 of the β-chain of hemoglobin. This mutant (sickle cell) hemoglobin remains in solution as long as oxygen concentration is high. However, at low oxygen concentrations, filamentous aggregates of the sickle cell hemoglobin precipitate causing the red blood cells to assume the characteristic sickle shape. It is from this phenomenon the disease derives its name. In addition to this primary effect, a number of other characters of the individuals homozygous for this gene are also affected. These effects include, among other things, 'tower skull', dialation of heart and heart failure, poor physical development, impaired mental function, pneumonia due to lung damage, rheumatism, paralysis, kidney damage and failure etc.

A number of other recessive genes produce marked and, often, detrimental pleiotropic effects. Consequently, they are referred to as ***syndromes***, *e.g., Hunters' syndrome, Huntington's chorea, cystic fibrosis* etc. Pleiotropic genes are known in many other organisms as well. In *Drosophila, white* eye gene (*w*) affects the shape of spermatheca in addition to producing white eyes. Several other genes of *Drosophila, e.g., vestigial* wings, *notched* wings, etc., are also pleiotropic.

Genes showing pleiotropy produce a single polypeptide just like other nonpleiotropic genes. But their polypeptides govern such a biochemical reaction, which is basic to many developmental events. As a result, the impairment of this function interferes with a number of developmental events, which in turn leads to the pleiotropic expression of the gene.

QUESTIONS

1. Describe the various systems of deriving gene symbols and discuss their merits and demerits.

2. Define dominance. Explain the various types of dominance with the help of suitable examples.
3. Classify the genes on the basis of their effect on survival of the individuals in which they are present in the appropriate genotype. Explain these classes of genes with the help of suitable examples.
4. Explain lethal gene action with the help of a suitable example.
5. Write short notes on the following:
   (*i*) wild type,
   (*ii*) supervital genes,
   (*iii*) lethal genes,
   (*iv*) codominance,
   (*v*) overdomiance,
   (*vi*) partial dominance,
   (*vii*) gene symbols derived from mutant phenotype,
   (*viii*) gene symbols for multiple alleles.

13

Gene Interaction and Epistasis (Modifications of Dihybrid Ratio)

13.1. INTRODUCTION

Mendel was lucky that in his studies each character was governed by a single gene. But the expression of many of the characters in almost all organisms is governed by two or more genes. These genes affect the development of the concerned characters in various ways. This gives rise to several modifications of the typical dihybrid (9 : 3 : 3 : 1) F_2 ratio. The phenomenon of two or more genes governing the development of a single character in such a way that they affect the expression of each other in various ways is known as ***gene interaction.***

Definition

- *Gene interaction* : A character governed by 2 or more genes, which affect the expression of each other.
- *Epistasis* : A gene affects the expression of another gene; both govern the same trait. Thus, *epistasis* describes the same events as *gene interaction.*

When one gene affects in any way the expression of another gene, the phenomenon is called ***epistasis.*** Thus *all cases of gene interaction are examples of epistasis.* But sometimes the term epistasis has been used to denote the masking effect of one gene on the expression of another gene. But such cases are best described as masking gene action (see later).

Gene interaction may involve two or more genes. But all the examples described below relate to interaction between two genes. Therefore, these cases are also referred to as the ***modifications of dihybrid ratio.***

13.2. TYPES OF GENE INTERACTION

There are several types of gene interaction. Some of the common gene interactions are as follows: (1) typical dihybrid ratio for a single trait (9 : 3 : 3 : 1), (2) duplicate gene action (15 :1), (3) complementary gene action (9 : 7), (4) supplementary gene action (9 : 3 : 4), (5) inhibitory gene action (13 : 3), (6) masking gene action (12 : 3 : 1), (7) polymeric gene action (9 : 6 : 1), and (8) additive gene action (1 : 4 : 6 : 4 : 1).

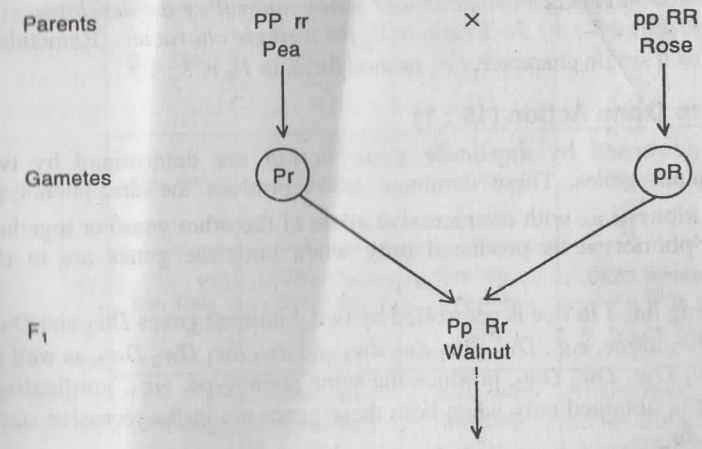
13.2.1. Typical Dihybrid Ratio for a Single Trait (9 : 3 : 3 : 1)

This type of gene interaction produces the typical dihybrid ratio of 9 : 3 : 3 : 1 in F_2 for a single character. Obviously, the concerned character is controlled by two

Gene Interaction: Type

1. *Dihybrid ratio*: F_2 ratio, 9 : 3 : 3 : 1.
2. *Duplicate gene action*: F_2 ratio, 15 : 1.
3. *Complementary gene action*: F_2 ratio, 9 : 7.
4. *Supplementary gene action*: F_2 ratio, 9 : 3 : 4.
5. *Inhibitory gene action*: F_2 ratio, 13 : 3.
6. *Masking gene action*: F_2 ratio, 12 : 3 : 1.
7. *Polymeric gene action*: F_2 ratio, 9 : 6 : 1.
8. *Additive gene action*: F_2 ratio, 1 : 4 : 6 : 4 : 1.

genes exhibiting complete dominance. The dominant alleles of each of the two genes produce separate forms of the character (phenotypes) when they are alone, *i.e.,* when the dominant allele of one gene is present with the homozygous recessive allele of the other gene. But when the dominant alleles of both the genes are present together, they produce a distinct phenotype. The homozygous recessive state at both the loci gives rise to yet another phenotype.



F_2

♀	PR	Pr	pR	pr
PR	PP RR Walnut	PP Rr Walnut	Pp RR Walnut	Pp Rr Walnut
Pr	PP Rr Walnut	PP rr Pea	Pp Rr Walnut	Pp rr Pea
pR	Pp RR Walnut	Rp Rr Walnut	pp RR Rose	pp Rr Rose
pr	Pp Rr Walnut	Pp rr Pea	pp Rr Rose	pp rr Single

Phenotypic ratio: 9 Walnut : 3 Pea : 3 Rose : 1 Single

Fig 13.1. A typical dihydrid (9 : 3 : 3 : 1) ratio in case of a single character, *viz.,* comb shape in poultry.

For example is chickens, comb shape is governed by two genes, *p* and *r*. The dominant allele of gene *p* alone produces pea comb, while that of gene *r* alone produces rose comb. Thus the genotype *PP rr* produces pea comb, while *pp RR* gives rise to rose comb. But when both the dominant genes *P* and *R* are present together, *e.g.*, *PP RR*, such individuals have walnut comb. Recessive condition at both the loci, *viz.*, *pp rr*, gives rise to a distinct comb shape called single.

When a breed of poultry homozygous for pea comb (*PP rr*) is crossed with another breed homozygous for rose comb (*pp RR*), the F_1 (*Pp Rr*) individuals have walnut comb as they have dominant alleles of both the genes *p* and *r*. Segregation for the two genes produces 16 possible zygotic combinations in F_2 (Fig. 13.1). Nine of these combinations have at least one *P* and one *R* allele. Consequently, these progeny have walnut comb. Three zygotes have one or two *P* alleles, but are homozygous *rr*. As a result, they have pea comb. Three other zygotes are homozygous for *p* (*pp*) but have one or two *R* alleles. Therefore, these individuals develop rose comb. The remaining one individual has the genotype *pp rr*; it develops single comb since it is homozygous recessive for both the genes *p* and *r*.

Thus an *interaction between two dominant genes controlling the development of a single trait generates a 9 : 3 : 3 : 1 ratio in* F_2 *for a single character.* (Remember, the typical ratio for a single character, *i.e.*, monohybrid, in F_2 is 3 : 1.)

13.2.2. Duplicate Gene Action (15 : 1)

Characters governed by ***duplicate gene action*** are determined by two completely dominant genes. These dominant genes produce the same phenotype whether they are alone (*i.e.*, with the recessive allele of the other gene) or together. The contrasting phenotype is produced only when both the genes are in the homozygous recessive state.

The nonfloating habit in rice is controlled by two dominant genes Dw_1 and Dw_2. Genes Dw_1 and Dw_2 alone, *e.g.*, $Dw_1\ Dw_1\ dw_2\ dw_2$ and $dw_1\ dw_1\ Dw_2\ Dw_2$, as well as together, *e.g.*, $Dw_1\ Dw_1\ Dw_2\ Dw_2$, produce the same phenotype, *viz.*, nonfloating. The floating habit is obtained only when both these genes are in the recessive state, *viz.*, $dw_1\ dw_1\ dw_2\ dw_2$.

When a nonfloating rice strain with the genotype ($Dw_1\ Dw_1\ Dw_2\ Dw_2$) is crossed with a floating strain ($dw_1\ dw_1\ dw_2\ dw_2$), the F_1 ($Dw_1\ dw_1\ Dw_2\ dw_2$) is nonfloating. In the F_2 generation, on an average, 9 plants will have at least one dominant allele of both the genes, 3 plants will have at least one dominant allele of one of the two genes (but will be homozygous recessive for the other gene), and 3 other individuals will have at least one dominant allele of the other gene. All these 15 individuals will have nonfloating habit (Fig. 13.2). Only one of the 16 possible zygotic combinations will be homozygous recessive for both the genes, dw_1 and dw_2, *viz.*, the genotype $dw_1\ dw_1\ dw_2\ dw_2$; this plant will have the floating habit.

Thus in ***duplicate gene action,*** *the presence of a single dominant allele of any one of the two genes governing the trait produces the dominant phenotype (e.g.,* nonfloating habit in rice). *The recessive phenotype (e.g.,* floating habit in rice) *is produced only when both the genes are in the homozygous recessive state.* The genes that show duplicate interaction are called ***duplicate genes.***

A similar gene interaction governs the fruit shape in *Capsella*, nodulation in

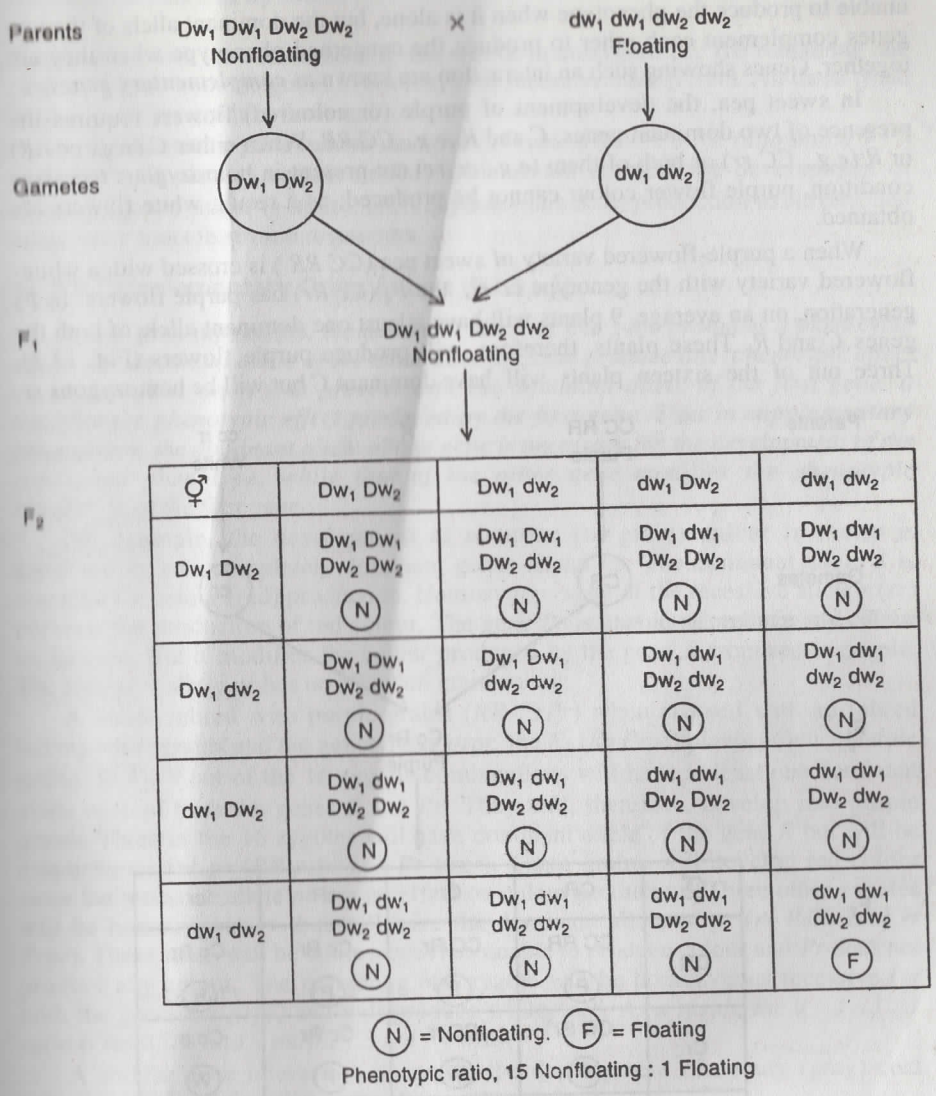


Fig. 13.2. Duplicate gene interaction in the development of floating habit in rice giving a 15 : 1 ratio in F_2.

groundnut and certain traits in many other organisms. This gene interaction produces 15 : 1 ratio in F_2.

13.2.3. Complementary Gene Action (9 : 7)

In this type of gene interaction, *the production of one phenotype requires the presence of dominant alleles of both the genes controlling the character. When any one of the two or both the genes are present in the homozygous recessive state, the contrasting phenotype is produced.* Thus any one of the two dominant genes is

unable to produce the phenotype when it is alone, but the dominant allels of the two genes complement each other to produce the concerned phenotype when they are together. Genes showing such an interaction are known as ***complementary genes***.

In sweet pea, the development of purple (or coloured) flowers requires the presence of two dominant genes, *C* and *R*, *e.g., CC RR*. When either *C (e.g., cc RR)* or *R (e.g., CC rr)* or both of them (*e.g., cc rr*) are present in homozygous recessive condition, purple flower colour cannot be produced; as a result, white flowers are obtained.

When a purple-flowered variety of sweet pea (*CC RR*) is crossed with a white-flowered variety with the genotype *cc rr,* the F_1 (*Cc Rr*) has purple flowers. In F_2 generation, on an average, 9 plants will have atleast one dominant allele of both the genes *C* and *R*. These plants, therefore, will produce purple flowers (Fig. 13.3). Three out of the sixteen plants will have dominant *C* but will be homozygous *rr*.

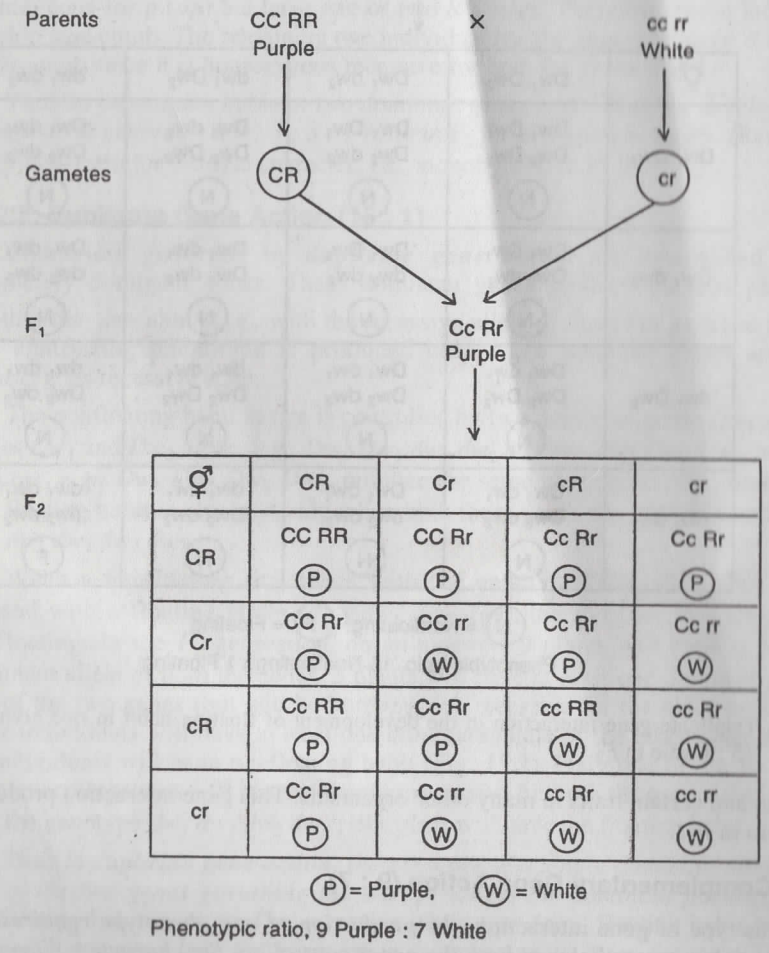


Fig. 13.3. Complementary gene interaction in the development of flower colour in sweet pea giving rise to a 9 : 7 ratio in F_2.

Three others will have dominant *R* but will be homozygous *cc*. The remaining one plant will have both the genes in homozygous recessive state (*cc rr*). All these (total 7) plants will have white flowers.

Complementary gene action *modifies the typical 9 : 3 : 3 : 1 ratio into a 9 : 7 ratio in* F_2. A similar complementary gene action controls the development of aleurone (grain) colour in maize, HCN (hydrocyanic acid) production in clovers, and many other traits in several organisms.

13.2.4. Supplementary Gene Action (9 : 3 : 4)

In this gene interaction, *the dominant allele of one gene produces a phenotypic effect. The dominant allele of the other gene does not produce any phenotypic effect on its own. But when it is present with the dominant allele of the first gene, it modifies the phenotypic effect produced by the first gene. Thus in* ***supplementary gene action****, the dominant allele of one gene is necessary for the development of the concerned phenotype, while that of the other gene modifies the phenotypic expression of the first gene.*

For example, the development of aleurone (or grain) colour in maize is governed by two completely dominant genes *R* and *Pr*. The dominant allele *R* is essential for colour (red) production. Homozygous state of the recessive allele *r* (*rr*) prevents the production of red colour. The gene *Pr* is unable to produce any colour on its own. But it modifies the colour produced by the gene *R* from red to purple. The recessive allele *pr* has no effect on grain colour.

A maize inbred with purple grains (*RR PrPr*) when crossed with an inbred having white grains and the genotype *rr prpr,* the F_1 (*Rr Prpr*) plants produce purple grains. In F_2, 9 out of the 16 zygotic combinations will have at least one dominant allele each of both the genes *R* and *Pr*. They will, therefore, develop into purple grains. Three of the 16 zygotes will have dominant allele of the gene *R* but will be homozygous for *pr* (*RR prpr* and *Rr prpr*). These grains will develop red colour since the recessive allele *pr* has no effect on colour production. Three other zygotes will be homozygous *rr* but will have the dominant allele of *pr* (*rr PrPr and rr Prpr*). These seeds will be white since *rr* is unable to produce colour and *Pr* does not produce any colour. The remaining one zygote will be homozygous recessive for both the genes (*rr prpr*) and will produce white seed. *As a result, the 9 : 3 : 3 : 1 ratio is modified into 9 purple : 3 red : 4 white.*

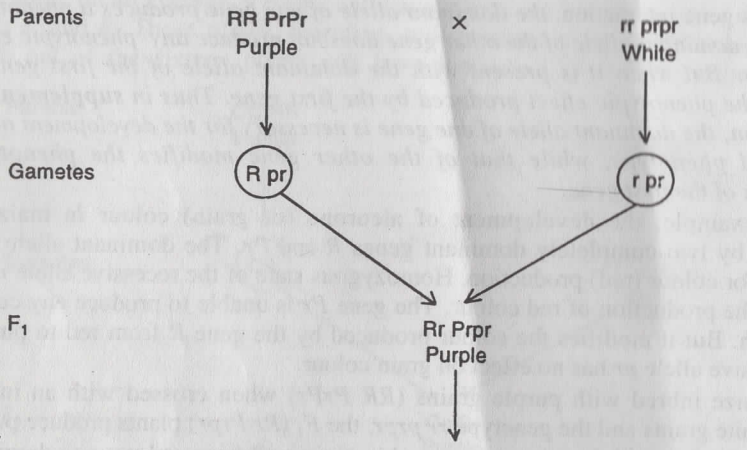
A similar gene interaction occurs in the development of agouty (gray) coat colour in mice and for other traits of several organisms.

13.2.5. Inhibitory Gene Action (13 : 3)

In ***inhibitory gene action,*** *one dominant gene produces the concerned phenotype or the character, while its recessive allele produces the contrasting phenotype. The second dominant gene has no effect of its own on the character in question, but it stops the expression of the dominant allele of the first gene. As a result, when the two dominant genes are present together, they produce the same phenotype as that produced by the recessive homozygote of the first gene. The recessive allele of the second gene does not affect the development of the character in any way.* Thus in ***inhibitory gene action,*** one dominant gene is capable of producing a phenotype only if its expression is not prevented by another dominant gene. The gene that stops the expression of another gene is called ***inhibitory gene***

and is denoted by *I*. Thus the 9 : 3 : 3 : 1 F_2 ratio is modified into 13 : 3 ratio in this case.

An example of inhibitory gene interaction occurs in the development of aleurone colour in maize. The dominant gene *R* produces red colour, while its recessive allele *r* produces no colour. Another dominant gene *I* does not produce any colour by itself, but it prevents colour production by gene *R* when both *I* and *R* are present together. The recessive allele *i* does not affect the colour production in any way. As a result, red colour (or any other colour) in aleurone is produced only when *R* is present with the homozygous recessive allele of the inhibitory gene *i* (*e.g., RR ii*).



F2

♀ / ♂	R Pr	R pr	r Pr	r pr
R Pr	RR PrPr (P)	RR Prpr (P)	Rr PrPr (P)	Rr Prpr (P)
R pr	RR Prpr (P)	RR prpr (R)	Rr Prpr (P)	Rr prpr (R)
r Pr	Rr PrPr (P)	Rr Prpr (P)	rr PrPr (W)	rr Prpr (W)
r pr	Rr Prpr (P)	Rr prpr (R)	rr Prpr (W)	rr prpr (W)

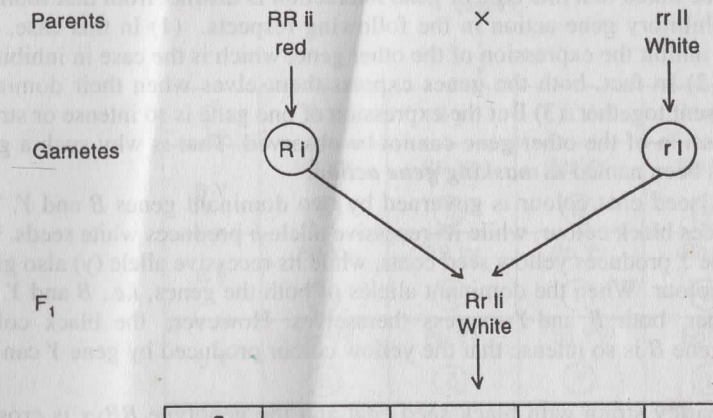
(P) = Purple, (R) = Red, (W) = White

Phenotypic ratio, 9 Purple : 3 Red : 4 White

Fig. 13.4. Supplementary gene action in the development of grain (aleurone) colour in maize producing 9 : 3 : 4 ratio in F_2.

When a maize inbred with red aleurone (*RR ii*) is crossed with another inbred having the genotype *rr II* and white grains, the F_1 (*Rr Ii*) has white grains. This is because gene *I* stops the colour production by *R*. In F_2 generation, on an average,

nine out of sixteen seeds will have at least one dominant allele of both *R* and *I* (Fig. 13.5). These seeds will develop no colour due to the inhibitory action of *I* on colour production by *R*, and will be white. Three seeds (out of 16) will have the dominant allele *R*, either in homozygous (*RR ii*) or in heterozygous state (*Rr ii*), but will be homozygous for the recessive allele of *i*. These seeds will develop red colour since the gene *R* is able to function in the absence of the dominant allele of *I*. Three other seeds will be homozygous *rr*, but will have one or two dominant alleles of *I*. The remaining one seed will be homozygous recessive for both the genes (*rr ii*). These seeds will be white simply because the colour producing gene is in the recessive state (*rr*).



F2

♀ / ♂	R I	R i	r I	r i
R I	RR II (W)	RR Ii (W)	Rr II (W)	Rr Ii (W)
R i	RR Ii (W)	RR ii (R)	Rr Ii (W)	Rr ii (R)
r I	Rr II (W)	Rr Ii (W)	rr II (W)	rr Ii (W)
r i	Rr Ii (W)	Rr ii (R)	rr Ii (W)	rr ii (W)

(W) = White, (R) = Red

Phenotypic ratio, 13 White : 3 Red

Fig. 13.5. Inhibitory gene action in the control of seed colour in maize, modifying the dihybrid F_2 ratio (9 : 3 : 3 : 1) into 13 : 3.

Thus in ***inhibitory gene action,*** *one dominant inhibitory gene prevents the expression of another dominant gene. This modifies the typical 9 : 3 : 3 : 1* F_2 *ratio into a 13 : 3 ratio and leads to an excess of the recessive phenotype (white grain) produced by the functional gene.*

Other examples of inhibitory gene action are the development of plumage (feather) colour in poultry, the production of anthocyanin pigment in rice and other traits in several other organisms.

13.2.6. Masking Gene Action (12: 3 : 1)

In this interaction, dominant alleles of the two genes affecting a character produce distinct phenotypes when they are alone. But when dominant alleles of both the genes are present together, the expression of dominant allele of one gene masks the expression of the other. When both the genes are present in recessive state, a different phenotype is produced

It should be noted that this type of gene interaction is distinct from that found in the case of inhibitory gene action in the following respects. (1) In this case, one gene does not inhibit the expression of the other gene, which is the case in inhibitory gene action. (2) In fact, both the genes express themselves when their dominant alleles are present together. (3) But the expression of one gene is so intense or strong that the expression of the other gene cannot be observed. That is why such a gene interaction has been named as ***masking gene action***.

In barley, seed coat colour is governed by two dominant genes *B* and *Y*. The allele *B* produces black colour, while its recessive allele *b* produces white seeds. The dominant allele *Y* produces yellow seed coats, while its recessive allele (*y*) also gives rise to white colour. When the dominant alleles of both the genes, *i.e.*, *B* and *Y*, are present together, both *B* and *Y* express themselves. However, the black colour produced by gene *B* is so intense that the yellow colour produced by gene *Y* can not be detected.

When a barley strain with black seed coat and the genotype *BByy* is crossed with another strain having yellow seed coat *bbYY*, the F_1 (*BbYy*) seeds have black seed coat. In F_2, on an average 9 seeds will have at least one dominant allele of both the genes *B* and *Y*. These seeds will develop black seed coats since only the black colour can be detected. Three of the sixteen zygotic combinations will have a dominant allele of *B* but will be homozygous for *y* (*BByy, Bbyy*). These seeds will also have black seed coat due to the presence of *B*. In three other seeds, recessive allele *b* will be in homozygous state, while one or two dominant alleles of *Y* will be present (*bbYY, bbYy*). These seeds will have yellow seed coats due to *Y* since gene *B* is in the recessive state. The remaining one zygotic combination (*bbyy*) will be homozygous recessive for both the genes; therefore, it will develop white seed coat colour (Fig. 13.6).

As a consequence of ***masking gene action,*** *the 9 : 3 : 3 : 1* dihybrid F_2 *ratio is modified into a 12 : 3 : 1 ratio (12 black : 3 yellow : 1 white seed coat colour*). A similar interaction is known to govern seed colour in jowar, fruit colour in summer squash, and certain traits in many other organisms.

13.2.7. Polymeric Gene Action (9 : 6 : 1)

In polymeric gene action, the two completely dominant genes controlling a character produce the same phenotype when their dominant alleles are alone. But when dominant alleles of both the genes are present together, their phenotypic effect is enhanced as if the effect of the two genes were cumulative or additive.

In barley, two completely dominant genes *A* and *B* affect the length of awns (the thin needle like extensions of lemma). Dominant allele of the gene *A* or *B* alone

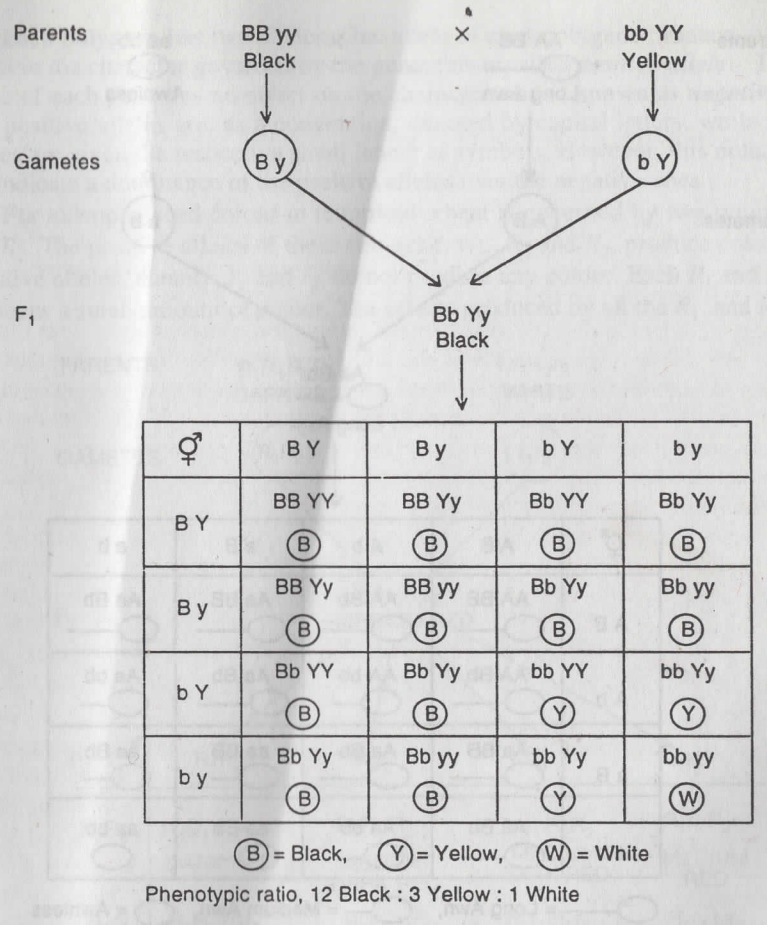


Fig. 13.6. Masking gene action in the control of seed coat colour in barely generating a 12 : 3 : 1 ratio in F_2

(*e.g., Aa bb, AA bb,* and *aa Bb, aa BB*, respectively) give rise to awns of medium length. Thus the phenotypic effect of *A* is the same as that of *B*. But when dominant alleles of both the genes *A* and *B* are present together (*e.g., Aa Bb*, etc.) they produce long awns. Individuals homozygous recessive for both these genes (*aa bb*) are awnless.

When a long-awned (*AA BB*) variety of barley is crossed with an awnless (*aa bb*) one, the F_1 (*Aa Bb*) has long awns due to the cumulative effects of the dominant alleles of both the genes *A* and *B*. In F_2 generation, on an average, nine plants (out of 16) will have dominant alleles of both the genes *A* and *B*. As a result, these plants will produce long awns. Three plants will have at least one dominant allele of *A* but will be homozygous for *b* (*AA bb, Aa bb*). Three others will be homozygous for *a* but will have at least one dominant allele of *B* (*aa BB, aa Bb*). All these six plants will have medium-sized awns since they have either *A* or *B*. The remaining one plant (out of the 16 combinations in F_2) will be homozygous recessive for both the genes (*aa bb*); it will, therefore, be awnless (Fig. 13.7).

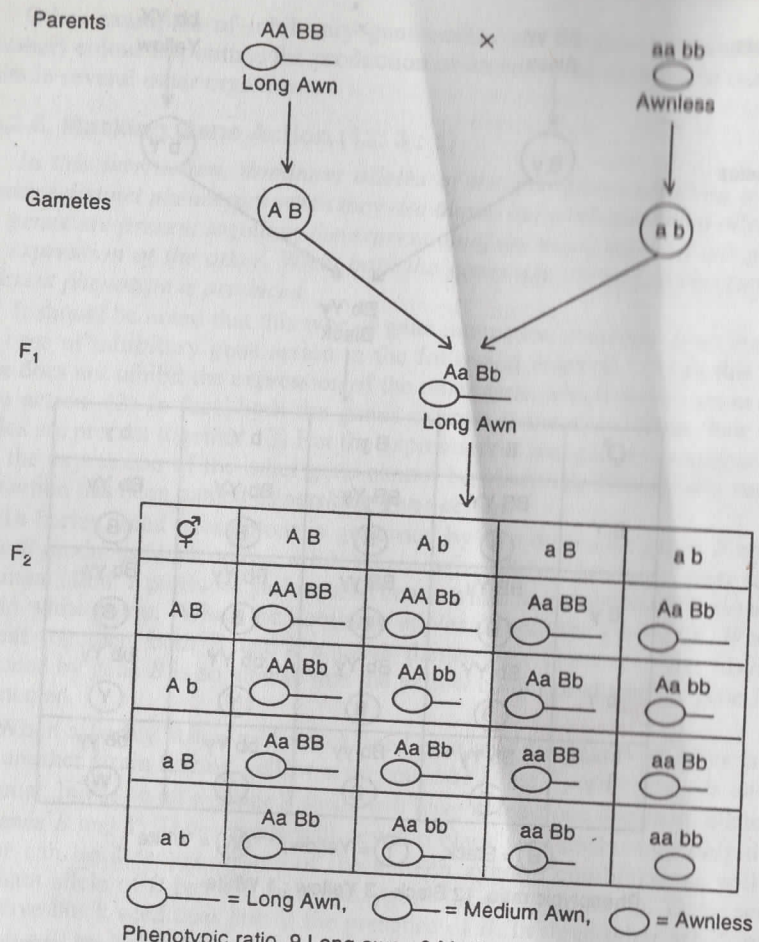


Fig. 13.7. Polymeric gene action determining awn development in barley; it gives rise to 9 : 6 : 1 ratio in F_2.

Thus ***polymeric gene interaction*** *changes the typical 9 : 3 : 3 : 1 dihybrid* F_2 *ratio into a 9 : 6 : 1 ratio.* Some other example of this interaction are also known, *e.g.*, fruit shape in summer squash etc.

13.2.8. Additive Gene Action (1 : 4 : 6 : 4 : 1)

In ***additive gene action,*** each positive allele of the two genes governing a trait produces equal and identical effect on the character. The effects of all the positive alleles present in an individual at the two loci are additive. As a result, the 9 : 3 : 3 : 1 dihybrid F_2 ratio is partitioned into the 1 : 4 : 6 : 4 : 1 ratio. This gene action is the basis for the ***multiple factor hypothesis*** of Nilsson-Ehle. Therefore, the genes showing additive gene action are called ***multiple factors*** or, more commonly, ***polygenes.***

Each polygene has two alleles. One allele of each polygene produces a positive effect in the character governed by the gene; this is called ***positive allele.*** The other allele of each gene has no effect on the character, and is known as ***negative allele.*** The positive alleles are, as a convention, denoted by capital letters, while negative alleles are given the respective small letters as symbols. However, this notation does not indicate a dominance of the positive alleles over the negative ones.

For example, seed colour in tetraploid wheat is governed by two polygenes R_1 and R_2. The positive alleles of these two gene, *viz.*, R_1 and R_2, produce colour. Their negative alleles, namely, r_1 and r_2, do not produce any colour. Each R_1 and R_2 allele produces a small amount of colour. The effects produced by all the R_1 and R_2 alleles

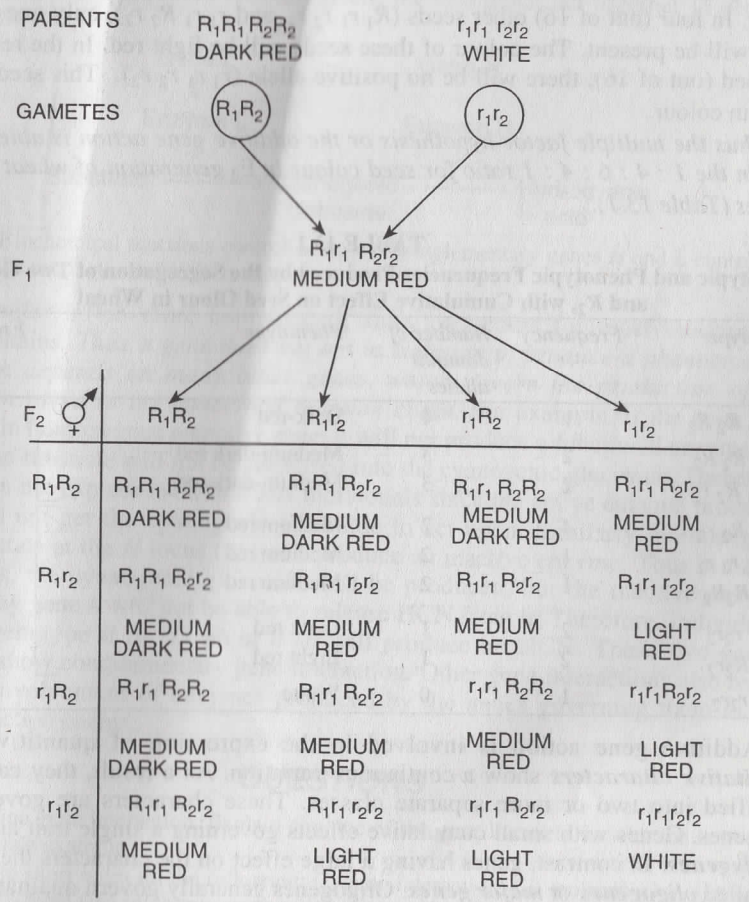


Fig. 13.8. Inheritance of seed colour in wheat; typical F_2 ratio is 1 : 4 : 6 : 4 : 1.

are additive. Therefore, the total intensity of colour produced depends on the total number of positive alleles of the two genes present. Thus $R_1 R_1 R_2 R_2$ produces dark red colour, $R_1R_1R_2r_2$ generates medium-dark red, $R_1r_1 R_2 r_2$, etc. yield medium red and so on.

When a variety having dark red ($R_1R_1\ R_2R_2$) seeds is crossed with a variety having white seeds ($r_1r_1\ r_2r_2$), the F_1 ($R_1r_1\ R_2r_2$) has medium red seeds. This is because the F_1 seeds have only two positive alleles (one R_1 and one R_2). In the F_2 generation (Fig. 13.8), one zygote out of sixteen will have four positive alleles ($R_1\ R_1\ R_2\ R_2$). This seed will be identical to one of the parents, and will have dark red colour. In four (out of 16) seeds, three positive alleles will be present ($R_1\ r_1\ R_2\ R_2$ and $R_1\ R_1\ R_2\ r_2$). These seeds will be lighter in colour than the seeds of the dark red coloured parent but darker than the F_1 seeds; they will be medium-dark red in colour. Six (out of 16) seeds will have only two positive alleles each ($R_1\ r_1\ R_2\ r_2$, $R_1\ R_1\ r_2\ r_2$, $r_1\ r_1\ R_2\ R_2$). They will be identical to the F_1 seeds and medium red in colour. In four (out of 16) other seeds ($R_1\ r_1\ r_2\ r_2$, and $r_1\ r_1\ R_2\ r_2$), only one positive allele will be present. The colour of these seeds will be light red. In the remaining one seed (out of 16), there will be no positive allele ($r_1\ r_1\ r_2\ r_2$). This seed will be white in colour.

Thus the multiple factor hypothesis or the additive gene action is able to fully explain the 1 : 4 : 6 : 4 : 1 ratio for seed colour in F_2 *generation of wheat and oat crosses (Table 13.1).*

TABLE 13.1

Genotypic and Phenotypic Frequencies Produced by the Segregation of Two Genes, R_1 and R_2, with Cumulative Effect on Seed Olour in Wheat

Genotype	*Frequency*	*Number of dominnt alleles*	*Phenotype*	*Frequency*
$R_1\ R_1\ R_2\ R_2$	1	4	Dark-red	1
$R_1\ r_1\ R_2\ R_2$	2	3	Medium-dark red	4
$R_1\ R_1\ R_2\ r_2$	2	3	Medium-dark red	
$R_1\ r_1\ R_2\ r_2$	4	2	Medium red	6
$R_1\ R_1\ r_2\ r_2$	1	2	Medium red	
$r_1\ r_1\ R_2R_2$	1	2	Medium red	
$R_1\ r_1\ r_2r_2$	2	1	Light red	4
$r_1\ r_1\ R_2r_2$	2	1	Light red	
$r_1\ r_1\ r_2r_2$	1	0	White	1

Additive gene action is involved in the expression of quantitive traits. ***Quatitative characters*** show a continuous variation. As a result, they can not be calssified into two or more separate classes. These characters are governed by polygenes. Genes with small cumulative effects governing a single trait are known as ***polygenes.*** In contrast, genes having a large effect on the characters they govern are called ***oligogenes*** or ***major genes***. Oligogenes generally govern qualitative traits with discrete variation. They permit classification of individuals into few discrete classes for the characters they govern. All the other cases of gene interaction described in this chapter are due to oligogenes.

13.3. MOLECULAR BASIS FOR GENE INTERACTION

It is customary to say that a gene (say *H* or *L*) produces a given character (say, HCN in clovers). But as a matter of fact, no gene produces a character directly by itself.

Genes generally produce their phenotypic effects by directing the synthesis of specific proteins. In most cases, these proteins act as enzymes. These enzymes catalyse specific biochemical reactions, which ultimately are responsible for the observed phenotypic effects. A biochemical reaction does not occur in isolation. Several biochemical reactions form a reaction chain in which the product of the previous reaction serves as the substrate for the next reaction (Fig. 13.9). For example, the production of HCN (hydrocyanic acid) from a substrate in clovers involves two biochemical reactions. The first reaction converts the substrate into a cyanogenic glucoside. The enzyme that catalyses this reaction is produced by gene *L*. The release of HCN from the cyanogenic glucoside is accomplished by another reaction, which is catalyzed by the enzyme produced by gene *H*. Clearly, neither gene *L* nor gene *H* produces HCN by itself; the two genes act together to produce HCN in clovers.

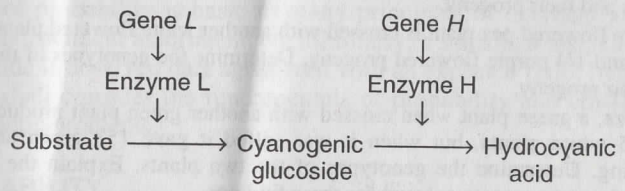


Fig. 13.9. Biochemical reactions controlled by the complementary genes *H* and *L* controlling the production of *HCN* in clovers.

A similar, often more complicated, situation occurs for most biochemical reaction chains. *Thus a gene does not act in isolation to produce a phenotype. Its expression depends on many other genes, which govern the production of the enzymes involved in the concerned reaction chain.* For example, if the gene *L* in clovers is in homozygous recessive state, it will not produce a functional enzyme. As a result, the substrate will not be converted into the cyanogenic glucoside. Therefore, gene *H* can not express itself in *ll HH* individuals since the active enzyme produced by *H* does not get the cyanogenic glucoside to act upon. Similarly, homozygous recessive state at the *H* locus (*hh*) will produce an inactive enzyme. Thus in *LL hh* individuals, the cyanogenic glucoside will be produced, but the inactive enzyme produced by gene *h* will not be able to release HCN from it. Therefore, individuals with the genotype *ll H-*, *L- hh* or *ll hh* will produce no HCN. These two genes, therefore, show complementary gene interaction. Other gene interactions also result from the involvement of enzymes produced by the genes governing them in the various reaction chains.

QUESTIONS

1. Define gene interaction. Explain any one of the gene interactions with the help of a suitable example.
2. Write short notes on the following: (*i*) gene interaction, (*ii*) molecular basis of gene interaction, (*iii*) inhibitory gene action, (*iv*) masking gene action, (*v*) duplicate gene action, (*vi*) complementary gene action, (*vii*) supplementary gene action, (*viii*) polymeric gene action, (*ix*) additive gene action .
3. In poultry, walnut comb shape is the result of interaction between two dominant genes *P* and *R*. Determine the comb shape of parents in the following crosses, the gametes produced by them, and the genotypes and phenotypes of their progeny.
   (*i*) *PP RR* × *Pp Rr* (*ii*) *Pp Rr* × *pp rr*

(iii) $Pp\ Rr \times Pp\ rr$ (iv) $Pp\ Rr \times pp\ Rr$
(v) $Pp\ Rr \times Pp\ RR$ (vi) $Pp\ Rr \times PP\ Rr$
(vii) $PP\ Rr \times pp\ rr$ (viii) $Pp\ RR \times pp\ rr$

4. A walnut combed hen was mated to a pea combed cock; it produced only one progeny, which had single comb. Determine the genotypes of the two parents and the genotypes and phenotypes of the progeny expected from matings between such birds.
5. In sweet peas, dominant genes *C* and *R* together produce coloured flowers. Determine the gametes, and genotypes and phenotypes of progeny obtained in the following crosses.
   (i) $Cc\ Rr \times cc\ rr$ (ii) $Cc\ Rr \times CC\ rr$
   (iii) $cc\ Rr \times Cc\ rr$ (iv) $Cc\ Rr \times cc\ Rr$
6. A white flowered sweet pea plant when crossed with one producing purple flowers produced 3/8 purple and 5/8 white progeny. Determine the genotypes of the two parents and their progeny.
7. A white flowered pea plant is crossed with another white flowered plant; it yields 3/4 white and 1/4 purple flowered progeny. Determine the genotypes of the two parents and their progeny.
8. In maize, a green plant when crossed with another green plant produced 225 green and 75 white plants, but when it was selfed it gave 153 green and 118 white offspring. Determine the genotypes of the two plants. Explain the type of gene interaction necessary to account for these findings.
9. In maize, gene *R* produces red aleurone colour, while gene *Pr*, when present together with *R*, produces purple colour. Determine the phenotypes of the parents, and the genotypes and phenotypes of the progeny produced from the following crosses.
   (i) $RR\ prpr \times Rr\ prpr$ (ii) $RR\ prpr \times Rr\ PrPr$
   (iii) $Rr\ Prpr \times Rr\ prpr$ (iv) $Rr\ Prpr \times rr\ Prpr$
   (v) $rr\ PrPr \times Rr\ Prpr$ (vi) $Rr\ Prpr \times Rr\ PrPr$
   (vii) $Rr\ Prpr \times rr\ prpr$
10. In mice, a dominant gene *A* affects the distribution within hairs of the black pigment produced by another dominant gene *C* so that *CC AA* genotype has gray hairs. Determine the phenotypes of the gametes produced by the parents in the following crosses, and the genotypes and phenotypes of the progeny produced by them.
    (i) $CC\ AA \times Cc\ Aa$ (ii) $Cc\ Aa \times cc\ aa$
    (iii) $Cc\ Aa \times Cc\ aa$ (iv) $Cc\ Aa \times cc\ Aa$
    (v) $Cc\ Aa \times Cc\ AA$ (vi) $Cc\ Aa \times CC\ Aa$
    (vii) $CC\ Aa \times cc\ aa$ (viii) $Cc\ AA \times cc\ aa$
11. In poultry, a dominant gene *I* inhibits plumage colour production by the dominant gene *C*. Determine the phenotypes of and the gametes produced by the parents and the genotypes and phenotypes of the progeny from the following crosses.
    (i) $Cc\ Ii \times Cc\ ii$ (ii) $Cc\ Ii \times cc\ ii$
    (iii) $Cc\ ii \times cc\ Ii$ (iv) $Cc\ II \times cc\ Ii$
    (v) $Cc\ Ii \times Cc\ II$
12. In summer squash, genes R_1 and R_2 produce round fruits when alone, but give rise to disc-shaped fruits when they are together. Determine the phenotypes of and the gametes produced by the parents, and the genotypes and phenotypes of the progeny from the following crosses.
    (i) $R_1r_1\ R_2r_2 \times R_1r_1\ r_2r_2$ (ii) $R_1r_1\ R_2r_2 \times r_1r_1\ r_2r_2$
    (iii) $R_1r_1\ R_2r_2 \times r_1r_1\ R_2r_2$ (iv) $R_1r_1\ R_2R_2 \times r_1r_1\ R_2\ r_2$
    (v) $R_1r_1\ R_2r_2 \times R_1r_1\ R_2R_2$

14

Probability and Chi-Square Test

14.1. INTRODUCTION

The concept of probability is basic to many principles of classical genetics, *e.g.*, segregation, independent assortment, etc. Similarly, a statistical test, chi-square test, is used to decide if observed data agree well with an expected ratio. In the following sections, we shall consider the fundamentals of probability and chi-square test as applied to genetic problems.

14.2. PROBABILITY

Simply defined, ***probability*** is the likelihood of occurrence of an event. In daily life, each one of us is often faced with a decision about the likelihood that an event will or will not occur. For example, before the commencement of examinations, each student is anxious about the likelihood of his doing well in the examinations. The likelihood of obtaining good marks in an examination may be influence by a number of factors, *e.g.*, the preparation by the student, the type of question paper (easy or hard), the attitude of examiner etc.

In contrast to such events, *there are a number of events whose occurrence depends entirely on chance.* For example, when a coin is tossed, either head or tail turns up. Similarly, a new-borne baby is either male or female. In both these cases, the events (turning up of head or tail in a toss, and the sex of a new-borne) occur purely because of chance. *The likelihood of occurrence of such events can be estimated and expressed in simple mathematical terms.* It is the probability of such events that we shall be discussing in this chapter.

Probability
- Likelihood of occurrence of an event.
- *Event* describes a happening, *i.e.*, occurrence of something.
  1. *Mutually exclusive events.*
  2. Independent events.

14.3. EVENTS

In simple terms, an ***event*** is the occurrence of something. The occurrence of head or tail in a toss is an event. Similarly, a new-borne baby being boy or girl is also an event. The events that occur purely due to chance are grouped into the following two categories on the basis of their relationship with each other: (*i*) mutually exclusive and (*ii*) independent events.

Mutually Exclusive Events

- Occurrence of one such event prevents the occurrence of others.
- P = 1/*n* (*n* = number of mutually exclusive events).
- Their probabilities add up to 1.

14.3.1. Mutually Exclusive Events

Mutually exclusive events are so related among themselves that the occurrence of one of them in a trial prevents the occurrence of the remaining events in the same trial. A ***trial*** refers to an opportunity provided for an event to occur. For example, tossing of a coin once constitutes one trial as it provides an opportunity for head or tail to occur. And in this trial, either head or tail can occur but not both. Thus in a trial, the occurrence of head excludes that of tail and *vice-versa*. Therefore, these events are mutually exclusive. *The probability of a mutually exclusive event is determined by a simple formula given below.*

$$P = 1/n$$

where, P = probability of a single mutually exclusive event, and n = the total number of mutually exclusive events.

In the case of toss of a coin, the total number of mutually exclusive events is two, *viz.*, head and tail. The probability of getting a head (or tail, for that matter) in a toss is, therefore, 1/2. The statement that the probability of getting head in a toss is 1/2 may be elaborated as follows. If a coin is tossed a sufficiently large (preferably, infinite) number of times, one-half of the times it will turn up head and the remaining one-half times it will give tail. Thus *the probability of an event is the average frequency of the event.* Conversely, *the average frequency of an event describes its probability.*

Many conditions are such that they can be satisfied by the occurrence of any one of two or more mutually exclusive events. The probability of such a condition being satisfied will be the total of probabilities of individual mutually exclusive events whose occurrence will satisfy that condition. For example, a dice has six faces, serially numbered 1 to 6, any one of which turns up in a throw. When a dice is cast, the probability of turning up of any number, say 4, will be 1/6 since $n = 6$. One may enquire about the probability of getting a number less than 5 in a throw. This condition will be fulfilled if any one of the numbers 1, 2, 3 and 4 is obtained. Therefore, the probability of getting a number less than five in a throw would be 4/6, which is the total of individual probabilities of getting 1, 2, 3 or 4 in a throw (=1/6 + 1/6 + 1/6 + 1/6).

Independent Events

- Such events do not affect the occurrence of each other.
- The probability of event *A* occurring with event *B* = probability of A × probability of B.

14.3.2. Independent Events

When two or more events are so related that the occurrence of one of the events does not affect the probability of occurrence of the remaining event(s), they are referred to as ***independent events.*** When two coins, say A and B, are tossed together or separately, the occurrence of head or tail in the first coin (A) does not affect the probability of occurrence of head or tail in the second coin (B). Thus the occurrence

of head or tail in coin A is independent of that in coin B; therefore, they are independent events.

The probability of occurrence of two or more independent events together is the product of individual probabilities of these events. For example, if two coins, *A* and B, are tossed together once, one may obtain any one of the following combinations of events: (1) head in both the coins; (2) head (H) in coin A, and tail (T) in coin B; (3) T in A, H in B; and (4) T in both the coins (Table 14.1). The probability (*P*) of obtaining H in both the coins will be the product of probabilities of getting H in coin A (1/2) and coin B (1/2) separately, that is, 1/2 × 1/2 = 1/4. Similarly, it can be shown that the probabilities of getting the remaining three combinations (H in A, T in B; T in A, H in B; and T in both) is 1/4 each (Table 14.1).

TABLE 14.1
Probability of Different Combinations of Head (H) and Tail (T) Obtained on Tossing Two Coins (A and B) Once Simultaneously or Separately.

S. No.	*Coin A*	*Coin B*	*Probability*
1	H	H	1/2 × 1/2 = 1/4
2	H	T	1/2 × 1/2 = 1/4
3	T	H	1/2 × 1/2= 1/4
4	T	T	1/2 × 1/2 = 1/4

The two combinations, H in A and T in B, and T in A and H in B, may be grouped together since they both have one H and one T. The only difference between these two combinations is the particular coin in which H or T was obtained. Thus the condition of obtaining one H and one T on tossing two coins once is satisfied by both these combinations. Therefore, the probability of getting one H and one T on tossing two coins will be the sum of the probabilities of these two combinations (= 1/4 + 1/4 = 1/2). This is because such combinations of independent events behave as mutually exclusive events. For example, when two coins are tossed once, one can get only one of the following four combinations: (1) H in both ; (2) H in A and T in B; (3) T in A and H in B; and (4) T in both.

Similarly, when three coins are tossed, there are eight possible combinations of H and T. These combinations are summarised below according to the order of the occurrence of H and T in the three coins.

HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

Each of these combinations will have the probability of 1/8 (= 1/2 × 1/2 × 1/2). These combinations can be further summarised on the basis of the numbers of H and T in them ignoring the sequence of occurrence of H and T; this yields the following four classes.

(1) HHH, (2) HHT, HTH, THH (3) HTT, THT, and TTH (4) TTT

These combinations may be represented as: (*i*) 1 (3H), (*ii*) 3 (2H 1T), (*iii*) 3 (1H 2T) and (*iv*) 1 (3T). Thus tossing of three coins together yields one combination of 3H, three different combinations of 2H 1T, three different combinations of 1H 2T, and one combination of 3T. Therefore, the probabilities of getting 3H, 2H 1T, 1H 2T and 3T on tossing of three coins together will be 1/8, 3/8, 3/8 and 1/8, respectively.

Binomial Expansion

- Expansion of $(a + b)^n$.
- Used to calculate probabilities of such cases, which have 2 mutually exclusive events.
- Coefficients of binomials obtained :
  1. from Pascal's triangle;
  2. by using formula $n\,!/s\,!\,t\,!$
- Probability of a combination of independent events = $\frac{n\,!}{s\,!\,t\,!} \times p^s\, q^t$.

14.4. BINOMIAL EXPANSION

The set of terms, along with their coefficients, obtained by expanding the general binomial $(a + b)^n$ is known as ***binomial expansion***. A complete expansion of the binomial $(a + b)^n$ involves two steps. The *first step* consists of the determination of various terms of the expansion. In the *second step*, the coefficients of these terms are determined. The first term obtained on simplification of $(a + b)^n$ is $a^n\, b^o$ or a^n (as $b^o = 1$). In the subsequent terms, the index of a sequentially decreases by 1 and that of b sequentially increases by 1 till it reaches $a^o\, b^n$ or b^n. Simplification of $(a + b)^n$ in this manner yields the following expansion.

$$(a + b)^n = a^n + a^{n-1}\, b^1 + a^{n-2}\, b^2 + \ldots\ldots\ldots + b^n$$

Each term of the above expansion has an appropriate coefficient. The coefficient for a term can be calculated by the following general formula.

$$\text{Coefficient} = n\,!/s\,!\,t\,!$$

where, n = the index of binomial, s = the index of a in the given term, t = the index of b in the term, and ! = factorial, *e.g.*, $n\,! = 1 \times 2 \times 3 \times 4 \times \ldots\ldots\ldots \times n$; $s\,! = 1 \times 2 \times 3 \times \ldots\ldots\ldots \times s$; and $t\,! = 1 \times 2 \times 3 \times \ldots\ldots\ldots \times t$ (in other words, $3\,! = 1 \times 2 \times 3$; $4\,! = 1 \times 2 \times 3 \times 4$; $5\,! \; 1 \times 2 \times 3 \times 4 \times 5$, etc.). The set of terms in the above expansion, when written along with the appropriate coefficients, constitutes the complete binomial expansion.

14.4.1. Pascal's Triangle

A relatively simple method for determining the set of coefficients for a binomial expansion is by using Pascal's triangle. This triangle was first constructed by Pascal, and is actually a set of figures written in a triangular shape. The Pascal's triangle is constructed by writing 1 at the top; this is the first column and the first row of the triangle. The subsequent figures are obtained according to the following rule: the number at a given position in the triangle will be the sum of two numbers, the number written immediately above this position and that written to the left-of-the-above this position. As per this rule, the next number can be written in the second row and column one only; this number will be 1 since above this position (in row one column one) is 1 and on the left-of-the-above (row one column zero) is 0. The third number of the triangle can be written in the second column of the second row; this number also will be 1 since the number above this position (row one column two) is 0 and that to the left-of-the-above (row one column one) is 1. The next number can be written only in column one of the row three; this number will be 1 since the number just above this position (row two column one) is 1 and that to the left-of-the-above (row two column zero) is 0. The second number in the third row (column two) will be 2 as the number above this position (row two column two) is 1 and that to the left-of-the-above (row two column one) is also 1. In this manner, one may write the set of numbers in as many rows as one wishes.

The sets of numbers in the Pascal's triangle have several properties. (1) The sum of number in a row is twice the sum of those in the preceding row, but only half of the sum of those in the following row. (2) The set of numbers in a row is symmetrical from the midpoint, *i.e.*, the numbers on one side of the midpoint are mirror images of those on the other side. (3) *The second number of a row (in the column two of a row) is the same as the index of binomial for whose terms the numbers of this row serve as coefficients.* For example, the number in column two of the sixth row of the triangle is 5. The set of numbers in this row (1, 5, 10, 10, 5, 1), therefore, are coefficients for the terms of binomial $(a + b)^5$. (4) *The sequence of numbers in a row is the same as that of terms of the binomial expanded in the manner explained earlier.*

Index of binomial	*Coefficients*										*Total*
0	1										1
1	1	1									2
2	1	2	1								4
3	1	3	3	1							8
4	1	4	6	4	1						16
5	1	5	10	10	5	1					32
6	1	6	15	20	15	6	1				64
7	1	7	21	35	35	21	7	1			128
8	1	8	28	56	70	56	28	8	1		256
9	1	9	36	84	126	126	84	36	9	1	512
10	1	10	45	120	210	252	210	120	45	10	1024

Fig. 14.1. Pascal's triangle.

14.4.2. Computation of Probabilities Using Binomial Expansion

The probabilities of different combinations of independent events can be computed easily by using the appropriate binomial expansion. *The binomial expansion appropriate for a problem has the same index as the total number of independent events in the problem.* However, *binomial expansion can be applied only if there are only two mutually exclusive events.* One of the two mutually exclusive events is represented by a, while the other is represented by b. The indices of a and b in a term of the binomial are the same as the numbers in a given combination of the mutually exclusive events they represent. The probability of a given combination of independent events is obtained by simplifying that term of the binomial, which is appropriate for that combination.

The probability of occurrence of a combination of independent events is computed as follows.

1. ***Selection of the binomial*** appropriate for the problem. *The appropriate binomial has the same index as the number of independent events in the problem.*
2. ***Expansion of the binomial*** along with coefficients of the various terms of binomial

3. ***Selection of the appropriate term.*** *In the appropriate term, the index of* a *is equal to the number of one mutually exclusive event in the combination, while that of* b *is equal to the number of the other mutually exclusive event.*
4. ***Simplification of the selected term.*** *For simplification,* a *and* b *are replaced by the probabilities of the mutually exclusive events they represent.*

14.4.2.1. Example 1

If a coin is tossed 5 times (or 5 coins are tossed together), determine the probability of obtaining 3H and 2T.

1. The binomial appropriate for this problem is $(a + b)^5$ since the number of independent events is 5.
2. The binomial $(a + b)^5$ is expanded as follows.
$$(a + b)^5 = a^5 + a^4b + a^3b^2 + a^2b^3 + ab^4 + b^5$$
The coefficients for this expansion are available from the sixth row of the Pascal's triangle as this row has the number 5 in its second column. These numbers (1, 5, 10, 10, 5, 1) are in the same sequence as the terms to which they apply in the expanded binomial. To complete the binomial, the coefficients are written along with the appropriate terms of the binomial.
$$= 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5$$
$$= a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$$
3. Let a represent H (head) and b represent T (tail). The index of a in the term will represent the number of H in the appropriate combination of H and T, while that of b will be the number of T in the combination. *Therefore, the term appropriate for the combination* 3H 2T *will be* $10a^3b^2$.
4. The probability of getting 3H and 2T in five tosses of a coin will be obtained by simplifying the term $10a^3b^2$. For simplifying a term, a is replaced by the probability of H (=1/2), while b is replaced by that of T (=1/2). The term $10a^3b^2$ is simplified as follows.

Probability of 3H 2T = $10a^3b^2$
$= 10(1/2)^3\ (1/2)^2$ (on replacing a by the probability of H and b by that of T)
$= 10 \times 1/8 \times 1/4$
$= 10/32$

14.4.2.2. Example 2

A number of conditions may be defined in such a way that each of them can be satisfied by more than one combination of the independent events. For example, one may ask about the probability of getting at least 3H if a coin is tossed five times. This condition will be satisfied by any one of the following combinations of H and T since each of them has at least 3H: 3H 2T, 4H IT and 5H. Some other similar conditions are listed below: (*i*) more than 3H, (*ii*) less than 3H, (*iii*) not more than 3H etc. in 5 (or any other number of) tosses. The probability of such conditions being satisfied is determined as follows.

1. Determination of the combinations of independent events that satisfy the specified condition. This is generally very simple as explained above in the case of at least 3H in 5 tosses.
2. Computation of the probabilities of the different combinations.

3. Summation of the probabilities of these combinations yields the probability of the condition being satisfied; the procedure is explained below.

1. ***Probability of Getting At Least 3H in 5 Tosses.*** The probability is computed as follows.
   (*a*) This condition is satisfied by the following combinations of H and T since each of them has at least 3H.
   3H 2T, 4H 1T, 5H
   (*b*) The terms of binomial appropriate for these combinations are: $10a^3b^2$, $5a^4b$, and a^5, respectively.
   Simplification of these terms yields,
   $10a^3b^2$, $= 10(1/2)^3\ (1/2)^2 = 10/32$
   $5a^4b$ $= 5(1/2)^4\ (1/2) = 5/32$
   a^5 $= (1/2)^5 = 1/32$
   (*c*) Sum of the probabilities of the above 3 combinations will give the probability of occurrence of least 3H in 5 tosses. The summation yields, 10/32 + 5/32 + 1/32 = 16/32 or 1/2. Therefore, the probability of getting at least 3H in 5 tosses is 16/32 or 1/2.

2. ***Probability of Getting More Than 3H in 5 Tosses.*** The probability of occurrence of more than 3H in 5 tosses is estimated as follows.
   (*a*) This condition is satisfied by the following combinations of H and T: 4H 1T and 5H.
   (*b*) Probabilities of these combinations will be obtained by a simplification of the terms $5a^4b$ and a^5, which will yield 5/32 and 1/32, respectively.
   (*c*) Sum of these two values is 6/32. Therefore, the probability of getting less than 3H in 5 tosses is 6/32.

3. ***Probability of Getting Less Than 3H in 5 Tosses.*** This is calculated as follows:
   (*a*) This condition is satisfied by the following combinations of H and T: 2H 3T, 1H 4T and 5T.
   (*b*) Probabilities of these combinations may be obtained by solving the terms $10a^2b^3$, $5ab^4$ and b^5, respectively; this yields 10/32, 5/32 and 1/32, respectively.
   (*c*) Sum of these values is 16/32. Thus the probability of getting less than 3H in 5 tosses is 16/32 or 1/2.

4. ***Probability of Getting Not More Than 3H in 5 Tosses.*** The computation of probability is done as follows.
   (*a*) This condition is satisfied by the following combinations of H and T since the number of H in them is 3 or less: 3H 2T, 2H 3T, 1H 4T and 5T.
   (*b*) Probabilities of these combinations of H and T are derived by simplifying the following terms: $10a^3b^2$, $10a^2b^3$, $5ab^4$, and b^5, respectively. This yields, 10/32, 10/32, 5/32 and 1/32, respectively.
   (*c*) Sum of these values is 26/32. The probability of obtaining not more than 3H in 5 tosses, therefore, is 26/32.

14.5. FORMULA FOR DETERMINATION OF PROBABILITY OF A COMBINATION OF INDEPENDENT EVENTS

The probability of a given combination of independent events can also be determined according to the following formula.

$$P = \frac{n!}{s!\,t!} \times p^s q^t$$

where, P = probability of the combination of independent events, n = total number of independent events, s = number of one of the two mutually exclusive events in the combination, t = number of the other mutually exclusive event in the combination, p = probability of the first of the two mutually exclusive events, q = probability of the other mutually exclusive event, and ! = factorial.

14.5.1. Example 3

If a coin is tossed 7 times, determine the probability of the combination 4H 3T. In this case, $n = 7$ (the total number of tosses), $s = 4$ (the number of H, one of the two mutually exclusive events), $t = 3$ (the number of T, the other mutually exclusive event), $p = 1/2$ (probability of H) and $q = 1/2$ (probability of T).

$$\text{Probability of 4H 3T} = \frac{7!}{4!\,3!} \times (1/2)^4\,(1/2)^3$$

$$= \frac{7\times6\times5\times4\times3\times2\times1}{(4\times3\times2\times1)\,(3\times2\times1)} \times 1/16 \times 1/8$$

$$= 7 \times 5 \times 1/16 \times 1/8$$

$$= 35/128$$

The same probability is obtained when this problem is solved by using the binomial expansion. The appropriate binomial for this problem will be $(a + b)^7$. The expansion of this binomial yields, $(a + b)^7 = a^7 + 7a^6b + 21a^5b^2 + 35a^4b^3 + 35a^3b^4 + 21a^2b^5 + 7ab^6 + b^7$

The term appropriate for the combination 4H 3T is $35a^4b^3$. Simplification of this term yields, $35(1/2)^4\,(1/2)^3 = 35/128$, which is the same as that obtained by the formula given above.

In the example described above, the probabilities of the two mutually exclusive events, *viz.*, H and T, are equal, *i.e.*, 1/2. The binomial expansion as well as the above formula are applied in the same manner to the solution of those problems in which the two mutually exclusive events do not have equal probabilities. An example of such a problem is the appearance of dominant and recessive characters in the F_2 generation of a monohybrid cross.

14.6. CHI-SQUARE TEST

We get ***enumeration data*** by classifying the individuals of a sample into few distinct classes having contrasting forms of a character and then counting the number of individuals in each class. An example of such data is the number of round and wrinkled seeds in the F_2 generation from an appropriate cross in pea. Obviously, such data are divisible into few clear-cut classes and consist only of whole numbers (integers). *Enumeration data pertain to qualitative traits of organisms: such traits are little affected by the environment and permit the classification of individuals into few distinct groups.*

14.7. POPULATION AND SAMPLE

Any biological study is based on a limited number of individuals selected for the purpose; such a group of individuals is known as ***sample***. A sample is a small part of a real or imaginary population. A ***population*** consists of all the possible observations of a kind, and is generally regarded as having infinite number of individuals or observations. For example, one may study the world human population to determine the frequency of different races and race-mixtures, the proportion of various religious and ethnic groups or of some other trait. In this case, all the human beings living on earth constitute the population; this population is real and finite, although very large. The experimenter would base his study on a small fraction of this population, which will constitute the sample. Similarly, Mendel studied 7,324 seeds in F_2 of round × wrinkled crosses for determining the inheritance of seed shape in pea. This group of 7,324 seeds is the sample drawn from an F_2 population, which can be obtained from an infinite number of crosses between round and wrinkled seeded pea plants. In this case, the population is imaginary or hypothetical and infinitely large.

χ^2-Test

- Used for enumeration data.
- Determines if the observed data agree with an expected ratio.
- Applied to original data, and not to ratios etc.
- Each class of data should have about 50 observations.

14.8. PURPOSE OF CHI-SQUARE TEST

Consideration of time, labour, money and practicability necessitate the study of samples drawn from populations. If these samples are random, *i.e.*, each individual of the population was given equal opportunity to be included in the sample, their properties will reflect those of the population from which they were drawn. As a result, various statistics, *e.g.*, mean, variance, etc., calculated from a random sample are estimates of those of the parent population. Thus mean, ratio, variance, etc. calculated from a sample are close, but seldom identical, to those of the parent population. For example, Mendel obtained 787 tall and 277 dwarf pea plants out of a total of 1,064 F_2 plants; this yields a ratio of 2.84 tall : 1 dwarf. According to the law of segregation (Chapter 9), the ratio is expected to be 3 tall : 1 dwarf. Thus the observed ratio 2.84 : 1 is close to, but slightly different from, the expected ratio of 3 : 1.

In such situations, the experimenter has to decide whether the observed ratio (*e.g.*, 2.84 : 1) is an estimate of the expected ratio (*e.g.*, 3 : 1) or it is really different from the latter. Mendel had to base this decision on his personal judgement. But it is essential that such decisions be based on an objective criterion. Such a criterion is furnished by the Chi-square (χ^2, pronounced as Ki-square) test. *The purpose of χ^2-test, therefore, is to decide if a set of observed data is according to an expected ratio or if it agrees well with an expected or theoretical distribution.*

14.9. REQUIREMENTS FOR χ^2-TEST

The χ^2-test is a relatively simple statistical test requiring a small amount of computation. The following conditions must be satisfied for a proper application of χ^2- test.

14.9.1. Enumeration Data

χ^2-test is generally applied to enumeration data only.

14.9.2. Expected Ratio

The numbers of observations in different classes obtained in the experiment are the ***observed frequencies*** of those classes. The *χ^2-test is applied to determine if these frequencies are in accordance with some specified ratio,* ***the expected ratio.*** *The expected ratio is generally obvious from or implied in the experiment (or the problem) itself.* For example, in F_2 generation of a cross segregating for one completely dominant character, a ratio of 3 : 1 is expected, while in that segregating for two completely dominant characters the expected ratio will be 9 : 3 :3 : 1. The expected ratio is used to compute the expected frequencies of different classes in the data. *The expected ratio is not an arbitrary ratio; it depends on the nature of the experiment in question, and the previous knowledge about the results from such experiments.*

14.9.3. Number of Observations

The larger the total number of observations in an experiment, the greater the reliability of the conclusions from the χ^2-test. Statisticians believe that each class of the data should have 50 or more observations, otherwise the conclusions from a χ^2-test will not be fully reliable. But some scientists feel that in most experiments if each class has five or more observations the results from chi-square test are acceptably reliable. If any class of data having only two classes has less than 50 observations, Yates correction is applied to increase the precision of chi-square test (see later).

Chi-square test is applicable to the original data itself, and not to be ratios or per cent frequencies computed from them. For example, the original numbers of 787 tall and 277 dwarf plants will be used for χ^2-test; the ratio of tall and dwarf plants (2.84 : 1) or their per cent frequencies (73% and 27%) cannot be used for the test.

14.10. PROCEDURE FOR χ^2 -TEST

The Chi-square test involves the following steps: (1) formulation of a null hypothesis, (2) computation of χ^2 value from observed data. (3) determination of the appropriate table value of χ^2, and (4) drawing of the conclusion.

14.10.1. Formulation of Null Hypothesis

Formulation of null hypothesis is the first step in a χ^2 text. The ***null hypothesis*** states that the observed data are in agreement with the expected ratio. In other words, deviations of the observed data from the expected ratio are not real, *i.e.*, they are due to chance only. For example, the null hypothesis for the data in Table 14.2 will be that the frequency of tall and dwarf plants is according to 3 : 1 ratio (the expected ratio in this case) or that the deviations of the observed data from the expected ratio of 3 : 1 are due to chance only.

Once the null hypothesis is formulated, the deviations of the observed from the expected frequencies of different classes of the data will be considered to be due to chance only. The distribution of χ^2 values obtained only due to chance deviations is theoretically known. Now the value of χ^2 obtained from the data can be compared with those occurring in the known theoretical distribution. In this way, the probability that the calculated value of χ^2 was due to chance only can be determined.

This allows the decision to be made if the calculated value of χ^2 is due to chance only.

χ^2-Test : Procedure

1. Formulate null hypothesis.
2. Estimate expected frequencies based on the expected ratio.
3. Determine the deviations of observed from expected frequencies.
   = (O – E).
4. Square the deviations and divide by E, *i.e.*, $(O - E)^2/E$.
5. Sum of all $(O - E)^2/E$ values is the value of χ^2.
6. Compare this value with the table value.

The calculated value of χ^2 can be compared with the values of the theoretical χ^2 only if the calculated value is assumed to be due to chance deviations only. *This is why it is essential to formulate the null hypothesis as the first step in a* χ^2-test.

14.10.2. Calculation of χ^2

Computation of χ^2 from the observed data is relatively simple; it is briefly described below.

1. The observed frequencies of different classes in a data are sequentially arranged as shown in Tables 14.2 and 14.3.
2. The expected frequencies of various classes are computed on the basis of expected ratio (stated in the null hypothesis) and the total number of observations in the observed data. For example, the total number of observations in the data of Table 14.2 is 1,064, and the expected ratio is 3 : 1. This expected ratio specifies that out of every 4 (= 3 + 1) plants, 3 will be tall and one will be dwarf. Therefore, out of the total of 1,064 plants, 266 (= 1 × 1,064/4) plants will be dwarf and 798 (=3 × 1,064/4) plants will be tall.

   Similarly, the total number of observations in the data of Table 14.3 is 556. The expected ratio in this case will be 9 : 3 : 3 : 1 since these data

TABLE 14.2
Chi-square Test on the Data of Mendel from F_2 Generation of Crosses between Tall and Dwarf Peas.
(Null hypothesis : the Observed Data are in the Ratio 3 : 1)

Description	*Number of plants in* F_2		*Total*
	Tall	*Dwarf*	
Observed frequency (*O*)	787	277	1,064
Expected frequency (*E*) (according to the 3 : 1 ratio)	798	266	1,064
Deviation (*O* – *E*)	(787 – 798) = –11	(277 – 266) = + 11	0
Square of the deviation $(O-E)^2$	$(-11)^2 = 121$	$(+11)^2 = 121$	
$(O - E)^2/E$	121/798 = 0.152	121/266 = 0.455	0.607
	$\chi^2 = \sum \frac{(O-E)^2}{E} = 0.607$		

pertain to the F_2 generation of a dihybrid cross (a cross segregating for two characters). According to the expected ratio, on an average, 9 seeds out of every 16 (= 9 + 3 + 3 + 1) F_2 seeds would be round yellow, 3 would be round green, 3 wrinkled yellow and 1 wrinkled green. Therefore, out of the total of 556 F_2 seeds, 312.75 (= 9 × 556/16) seeds are expected to be round yellow, 104.25 (= 3 × 556/16) round green, 104.25 (= 3 × 556/16) wrinkled yellow, and 34.75 (= 1 × 556/16) wrinkled green.

TABLE 14.3
Chi-square Test on the Data of Mendel from F_2 Generation of A Cross Segregating for Seed Shape and Seed Colour.
(Null hypothesis : The Observed Data Are in the Ratio 9 : 3 : 3 : 1)

Description	*Number of seeds in* F_2				
	Round yellow	*Round green*	*Wrinkled yellow*	*Wrinkled green*	*Total*
Observed frequency (*O*)	315	108	101	32	556
Expected frequency (*E*) (based on the 9:3:3:1 ratio)	312.75	104.25	104.25	34.75	556
Deviation (*O*–*E*)	+2.25	+3.75	–3.25	–2.75	0
Square of deviation $(O-E)^2$	5.06	14.06	10.57	7.46	
$(O-E)^2/E$	0.016	0.135	0.101	0.215	0.467

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 0.467$$

3. In the next step, deviation of the observed frequency from the expected one is determined for each class of the data. This is done by subtracting the expected frequency of a class from the observed frequency for that class. As a result, deviations have either + or – sign, and the sum of deviations for all the classes in a sample is ordinarily 0 (Tables 14.2, 14.3). When the expected frequencies have decimal fractions, the sum of deviations may not be 0 due to errors in rounding off of the decimal fractions. However, the sum of the deviations in such cases should be less than one.
4. Now the + and – signs of the deviations are removed by squaring the deviations.
5. The importance of a deviation depends not only on its own magnitude, but also on the value of the expected frequency from which this deviation has occurred. Therefore, the deviation squares are divided by the expected frequencies of the respective classes. All the values (quotients) thus obtained are based on the expected frequency of one and, therefore, are equal in importance.
6. Summation of all the quotients thus obtained yields the calculated value of chi-square. Thus *chi-square may be defined as the sum of the quotients obtained from division of squares of deviations of observed frequencies from the expected frequencies of various classes with the respective expected frequencies.* This may be expressed as follows.

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

where, Σ = summation, O = observed frequency of a class, and E = expected frequency of the class. Thus the calculated value of χ^2 from data in Table 14.2 is 0.607 and that from the data in Table 14.3 is 0.467.

χ^2-Test : Conclusion

- *Table value of* χ^2 : From χ^2 table :
  1. at 0.05 probability, and
  2. against (n – 1) digrees of freedom (where, n = number of classes in data).
- *Conclusion*:
  1. *Calculated value less than Table value:* ***null hypothesis accepted;*** *data agree with expected ratio.*
  2. *Calculated value equal to or higher than table value :* ***null hypothesis rejected;*** *data do not agree with expected ratio.*

14.10.3. Table Value of χ^2

The values of chi-square listed in the table (Table 14.4.) are entirely due to chance. These values serve as a point of reference while deciding whether a calculated value of χ^2 is due to real or chance deviations. The value of χ^2 depends on two variables: (1) degrees of freedom and (2) probability.

14.10.3.1. Degrees of Freedom

In a χ^2-test, the ***degrees of freedom*** (*d.f.*) is the number of classes of data that can be assigned arbitrary frequencies so that the total of the frequencies of all the classes is the same as that of the observed data. Therefore, the ***degrees of freedom is one less than the number of classes in the data; it can be represented as d.f. = n-1,*** where n = number of classes in the data.

The concept of degrees of freedom will become clearer, if we consider the data in Table 14.2. The data has two classes, tall and dwarf, and the total number of observations is 1,064. If you are asked to write any frequency (= number) in the two classes so that the total remains 1,064, can you really write *any number* in both the classes? At first, it would seem so. But if you think about it, you will find that you can fill an arbitrary frequency, say *X*, *in only one of the two classes.* The frequency of the remaining class will be the difference between 1,064 and the frequency chosen for the first class, *i.e., 1,064 –X.* Suppose you decide on 364 as the frequency of tall plants. Now the frequency of dwarf plants has to be 1,064 – 364, *i.e.,* 700. This is because if you write any other number, the total number of observations will differ from 1,064, which is not permissible. Once the total number of observations is specified, you are free to choose any number you desire as the frequency of only one of the two classes. Similarly, if the data has four classes (Table 14.3), you can assign arbitrary frequencies to only three of the four classes; the frequency of the fourth class will be the difference between the total number of observations in the data and the sum of the assigned frequencies of the three classes.

In the chi-square table (Table 14.4), degrees of freedom is written in the first column on the extreme left. This column has only whole numbers or integers. It may either be represented by the abbreviation *d.f.*, the greek letter υ (nue) or the formula (n – 1).

TABLE 14.4
The Maximum Values of Chi-square at Different Probabilities and Degrees of Freedom Obtained Purely Due to Chance

Degrees of freedom	*Probability (P)*					
(*d.f.*)	0.9	0.5	0.1	**0.05**	0.025	0.01
1	0.016	0.455	2.706	**3.841**	6.635	6.635
2	0.211	1.386	4.605	**5.991**	9.100	9.210
3	0.584	2.366	6.251	**7.815**	11.345	11.345
4	1.064	3.357	7.779	**9.488**	13.277	13.277
5	1.610	4.351	9.236	**11.070**	15.086	15.086
6	2.204	5.348	10.645	**12.592**	16.812	16.812
7	2.833	6.346	12.017	**14.067**	18.475	18.475
8	3.490	7.344	13.362	**15.507**	20.090	20.090
9	4.168	8.343	14.684	**16.919**	21.666	21.666
10	4.865	9.342	15.987	**18.307**	23.209	23.209

14.10.3.2. Probability Level

In the χ^2 table (Table 14.4), probability is listed in the topmost row and is often denoted by *P*. The value of probability is that fraction (from a very large number) of experiments with the specified *d.f.*, which will give as high or a higher value of χ^2 as that listed against the *d.f.* and under that probability due to chance only. For example, the value of χ^2 against 1 *d.f.* and under 0.05 *P* is 3.841. This means that out of a large number of similar experiments with 1 *d.f.*, 0.05 or 5% of the experiments will show a χ^2 value of 3.841 or more only due to chance. Conversely, in 0.95 or 95% of the experiments, a χ^2 value of 3.841 or of higher magnitude will be due to real deviation of the data from the expected ratio.

In most biological experiments, 0.05 probability is accepted as the standard probability level for decision making. In some situations, there may be sufficient justification for the experimenter to select a different probability level. But in such cases, the reason for selecting the probability level is clearly stated.

14.10.4. Arriving at A Conclusion

After the computation of χ^2 from the observed data, there are two ways in which a conclusion may be reached regarding the validity of null hypothesis. In the ***first method***, the value of χ^2 at 0.05 probability against the appropriate degrees of freedom is obtained from the χ^2 table. (1) *If the calculated value of χ^2 is less than this table value, the deviation of observed data from the expected frequencies is accepted to be purely due to chance. Therefore,* ***the null hypothesis is accepted****, and it is concluded that the observed data are in accordance with the expected ratio.* (2) *However, if the calculated value of χ^2 is equal to or greater than the table value of χ^2, the deviation of observed data from the expected frequencies is accepted to be real. In such a case,* ***the null hypothesis is rejected****, and it is concluded that the observed data are not according to the expected ratio.*

The calculated value of χ^2 from the data in Table 14.2 is 0.607. The χ^2 value from the χ^2 table at 0.05 probability and 1 *d.f.* is 3.841. The calculated value of χ^2 (0.607) is less than the table value (3.841), indicating the deviations to be due to chance. Therefore, the null hypothesis is accepted, and it is concluded that the data are in agreement with the 3 : 1 ratio. A similar conclusion is reached when the calculated value of χ^2 (0.467) from the data in Table 14.3 is compared with that from the χ^2 table (7.815) at 0.05 *P* and 3 *d.f.*

The ***alternative method*** for arriving at a conclusion is to determine the probability of the calculated value of χ^2. This is done by looking in the χ^2 - table against the *d.f.* appropriate for the data. The probability of χ^2 calculated from the data in Table 14.2 will be determined by comparing this value with those listed in the table against 1 *d.f.* This value (0.607) is more than 0.455 listed under 0.5 *P* and less than 2.706 under 0.1 *P*. The probability of the calculated value of χ^2 is, therefore, less than 0.5, and more than 0.1. Similarly, the probability for the value of χ^2 calculated from the data in Table 14.3 will be determined by comparing this value (0.467) with those listed in the table against 3 *d.f.* This value is less than 0.584, the χ^2 value listed under 0.9 *P*. Therefore, the probability of the calculated χ^2 in this case will be more than 0.9.

If the probability of calculated χ^2 is more than 0.05, ***the null hypothesis is accepted*** *and the deviations are regarded as due to chance only. But if the probability of calculated value of χ^2 is equal to or less than 0.05,* ***the null hypothesis is rejected*** *and the deviations are accepted as being real.*

Yates Correction
- Used for data having two classes.
- When number of observation in a class is < 50.
- The values of (*O* – *E*) for both the classes are reduced by 0.5 (signs of the values not considered).
- Rest procedure as usual.

14.11. YATES CORRECTION

When the data has only two classes and the frequency of one of the two classes is less that 50, Yates correction is applied to increase the precision of χ^2-test. In this correction, a deduction of 0.5 is made from the deviations of the two classes. While subtracting 0.5 the sign of the deviation is ignored. As a result, the two corrected deviation are 0.5 less than their original value. For example, deviation in the two classes of Table 14.5 are + 15 and – 15. Subtraction of 0.5 from + 15 yields 14.5. But a deduction of 0.5 from – 15. should yield 15.5 (=–15 – 0.5) if the sign of the deviation is taken into account. But since the sign of the deviation is ignored during

TABLE 14.5
Yates Correction for Enumeration Data with Two Classes. These Hypothetical Data Pertain to an F_2 Generation of a Cross for Seed Shape in Peas (Null Hypothesis: The Observed Data are in the Ratio 3 : 1)

Description	*Number of F_2 Seeds*		*Total*
	Round	*Wrinkled*	
Observed frequency (*O*)	165	35	200
Expected Frequency (*E*)	150	50	200
Deviation (*O* – *E*)	+ 15	– 15	0
Corrected deviation (= deviation less 0.5)	15 – 0.5 = 14.5	– (15 – 0.5) = 14.5	0
Square of corrected deviation	$14.5^2 = 210.25$	$-14.5^2 = 210.25$	
Square of corrected deviation / expected frequency	1.416	4.205	5.621
	$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 5.621$		

subtraction, the corrected deviation will be –14.5 [= –(15 – 0.5) in this case also. The mode of calculation of χ^2 from the corrected deviations and the drawing of conclusion remain the same as explained earlier. The calculated value of χ^2 from the data of Table 14.5 is 5.621, which is more than the table value of χ^2 at 0.05 *P* and 1 *d.f.* This indicates that the deviations of observed data from the expected frequencies are real. The null hypothesis is, therefore, rejected , and it is concluded that the observed data are not in 3 : 1 ratio.

QUESTIONS

1. Write short notes on the following:
   (*i*) probability,
   (*ii*) mutually exclusive events,
   (*iii*) independent events,
   (*iv*) χ^2-test, (*v*) null hypothesis,
   (*v*) degrees of freedom, and
   (*vi*) Yates correction.
2. A family has four children. What is the probability that:
   (*i*) all the four are boys,
   (*ii*) there are 2 girls and 2 boys,
   (*iii*) there are at least 2 boys,
   (*iv*) there are less than two girls, and
   (*v*) there are more than two boys?
3. A coin is tossed 6 times. Determine the probabilities of the following combinations of heads (H) and tails (T):
   (*i*) 4H 2T,
   (*ii*) 1H 5T,
   (*iii*) at least 3H,
   (*iv*) more than 4H,
   (*v*) less than 3H,
   (*vi*) not more than 3H and
   (*vii*) all H.
4. In a maternity home, five babies were borne on a single day. Estimate the probabilities of the following combinations of boys and girls in this case :
   (*i*) five boys,
   (*ii*) five girls,
   (*iii*) at least two boys,
   (*iv*) not more than 3 girls,
   (*v*) not more than 3 boys,
   (*vi*) less than 3 boys, and
   (*vii*) 3 boys and two girls.
   Will it make any difference if the five babies were borne on five different days ?
5. The Indian Cricket team is to play an eight test series in Australia. Captains of the two teams toss a coin before the start of every test match to decide who bats first. What is the probability that the Indian captain will win the toss in:
   (*i*) all the eight tests,
   (*ii*) five tests,
   (*iii*) one test only,
   (*iv*) at least four tests,
   (*v*) more than four tests,

   (*vi*) less than 3 tests,
   (*vii*) not more than 3 tests ?
6. A couple have five daughters. They are keen to have at least one son. Therefore, they decide to have another baby. What is the probability that the next baby will be a son ?
7. Six families have one child each. What are the probabilities of the following combinations of girls and boys in these six families ?
   (*i*) six boys,
   (*ii*) six girls,
   (*iii*) less than 2 girls,
   (*iv*) more that 3 girls, and
   (*v*) at least four boys.
8. Briefly describe the procedure for chi-square test. Discuss the applications of chi-square test.
9. The F_2 generation from a cross between albino (white) and black strains of mice produced 75 black and 33 albino progeny.
   (*i*) What is the expected ratio in this case ?
   (*ii*) State the null hypothesis.
   (*iii*) Determine the *d.f.*
   (*iv*) Carry out chi-square test to ascertain if the observed data agree with the expected ratio.
   (*v*) Describe the correction, if any, required in this case.
10. A coin was tossed 200 times. This yielded 90 heads and 110 tails. Conduct a χ^3-test to determine if the frequency of H and T in the observed data are in agreement with the expected ratio.
11. The census data of a village show that there are 527 males and 483 females in the village. Do these observations agree with the general belief that the frequencies of males and females in a population are equal ?

15

Environmental Effects on the Development of Characters

15.1. INTRODUCTION

It is generally said that a given gene produces or governs the development of a given character. This seems to imply that a gene produces a character irrespective of the other genes present in the organism as well as the environment to which the organism is exposed. However, the expression of a gene is affected in various ways by other genes present in the organism as well as by the environment.

Genotype, Phenotype and Environment

- *Genotype* : genetic constitution (*G*).
- *Phenotype* : external appearance for one or more traits (*P*).
- *Environment* : all factors other than the gene in question (*E*):

(*i*) External	(*ii*) Internal
— Temperature	— Age
— Light	— Sex
— Nutrition	— Substrate
— Humidity and water	— Tissue-specific expression
— Oxygen	— Genetic background.
— Maternal relations	
— Parasites	
— Symbionts	
— Population density	

- P = G + E (simply expressed)

15.2. GENOTYPE AND PHENOTYPE

The first clear-cut evidence that the development of a character is influenced by both genotype and environment was presented by Johannsen in 1911. Johannsen selected for seed size in a commercial seed sample of the Princess variety of rajma (*Phaseolus vulgaris*). Selection in the original seed lot was markedly effective, but it was ineffective in the individual plant progenies of the selected plants. Based on these findings, Johannsen proposed the universally accepted concepts of genotype and phenotype. ***Phenotype*** of an individual refers to its appearance in terms of its different characters; it is affected by both the genotype and the environment. The ***genotype*** of an individual denotes the alleles of the various genes present in the

individual. *Therefore, phenotype of an individual provides only a limited indication of its genotype. A reliable information on the genotype of an individual is furnished by the progeny of that individual.*

15.3. THE ENVIRONMENT

It is universally accepted that *the development of characters is under the joint control of genotype and environment.* Obviously, the expression of genes is affected by the environment. *Here,* ***environment*** may be defined as all the factors other than the gene under consideration. The various environmental factors may be grouped into two categories: (*i*) external, and (*ii*) internal.

15.4. THE EXTERNAL ENVIRONMENT

The ***external environment*** includes all the factors affecting an organism and existing outside the concerned organism. The external environment may be either (1) abiotic or (2) biotic. ***Abiotic*** factors are nonliving factors of the environment; some of the important abiotic factors are temperature, light, nutrients, moisture and water, and oxygen. In contrast, the ***biotic*** environment consists of living organisms that interact with an organism; these factors may be grouped as pathogens, insects and pests, symbionts and maternal relations.

Most genes are able to express themselves in a variety of environments, and do not need a specific environment for their expression. But many genes require a specific environment for development of the character governed by them. Such a character is known as ***threshold character,*** and the environment is called ***threshold.*** The threshold may be a specific temperature, nutrient, light or an organism. For example, a disease susceptible plant will develop the disease only when it is exposed to the concerned pathogen. In the absence of this pathogen it will be free from the disease and would appear as disease resistant. Many other examples of threshold characters are described in the following sections.

Temperature

- *Temperature sensitive* mutants; expressed at higher/lower temperatures, *e.g.,* TGMS.
- Heat shock proteins : act as chaperonins.
- *Chaperonins* : associate with unfolded proteins and prevent their denaturation.
- *Thermosensitive* : do not flower above/below threshold temperature.

15.4.1. Temperature

Temperature affects every phase of growth and development of an organism. The effect of temperature is mediated by its effect on the rate of enzyme activity. Genes produce their effects by producing specific proteins most of which are enzymes of various kinds. As a result, the expression of most of the genes is affected by temperature and other factors of the environment. But in some cases, the gene expression is dramatically affected by temperature. A recessive gene produces a peculiar (Himalayan) pattern of fur coloration in rabbits. Rabbits homozygous for this allele ($c^h c^h$) have black fur at the extremities of their body, *e.g.,* toes, snout, ears, and tail, while their main body in white. If the black hairs of Himalayan rabbits are shaved off, and the shaven area is kept warm, the new hairs that grow in these areas are white instead of black. Similarly, when white hairs are shaved off and the

shaven area is kept cool, the newly produced hairs are all black. Obviously, some enzyme governing melanin (the body and hair pigment) production is active at low temperature but inactive at a higher temperature. Such an enzyme is termed as ***temperature sensitive.*** In the himalayan rabbits, tyrosinase (the first enzyme involved in melanin production) is temperature sensitive in that it is inactive at higher temperatures.

In some animals, *e.g.*, turtles, crocodiles etc., the temperature of incubation of their eggs has a deciding effect on the sex of the individuals hatched from these eggs (Chapter 19, Section 5. 1. 3). *In plants temperature-sensitive genetic male sterility* (TGMS ; originally discovered in rice) is expressed only at higher temperatures; at temperatures lower than the threshold temperature, normal male fertility results. In case of rice TGMS line Pei-Ai 645, the threshold temperature is 23.3°C; at temperatures lower than this, plants are male fertile.

In many organisms, a set of genes are expressed only in response to a heat shock, *i.e.,* a brief exposure to a high temperature. These genes are called *hsp genes* (hsp, heat shock proteins). Heat shock proteins belong to a group of proteins called ***chaperonins***, which associate with unfolded proteins and prevent their incorrect folding, and the consequent denaturation.

When the expression of a gene depends on a specific temperature, it is called ***temperature sensitive***. Such temperature sensitive mutations are known in many organisms. Some varieties of crops like wheat and rice require a specific temperature for flowering; such varieties are known as ***thermosensitive.*** Thermosensitive varieties do not flower above or below the threshold range of temperature.

Light

- *Light sensitive*; genes expressed in response to light, *e.g., cab , rbcS.*
- Due to LRE (enhancer).
- *Photoperiod-sensitive*; expressed under specific photopreiod.
- *Photosensitive*; flower under long/short days.

15.4.2. Light

Sunlight is the source of all the energy available to the biological world through photosynthesis in green plants. Light affects plant growth and development by affecting the rate of photosynthesis, the concentration of some growth regulators like auxins, the development of chlorophyll and chloroplasts etc. Normal differentiation of chloroplasts and biosynthesis of chlorophyll depend on the presence of light even when plants have all the necessary genes for their development. Clearly, the biosynthesis of chlorophyll and chloroplast differentiation have a threshold requirement for light; such characters are known as ***light sensitive.***

Another example of a light sensitive trait is the *sun-red* character of maize. This character is governed by a single recessive gene. In a plant homozygous for the *sun-red* gene, the parts exposed to sunlight develop intense red colour, while those protected from sunlight remain green. This effect is produced by the blue-violet part of the visible spectrum.

In plants, expression of a kind of genetic male sterility is greatly affected by the photoperiod; this male sterility in known as ***photoperiod sensitive genetic male sterility*** (PGMS). In case of PGMS in rice, the *ms* (male sterility) gene is expressed

only when the day - length is Ca. 14 hr or longer provided the temperature is within the range of 23-29°C. The plants are normal male fertile if the days are of 13 hr or shorter.

Many plant genes are naturally expressed in response to light, *e.g.,* genes *rbcS* (encodes the smaller subunit of enzyme RUBISCO, ribulose- 1,5 - bisphosphate carboxylase/oxygenas), *cab*1 (encodes chlorophyll *a* / *b* binding protein) etc. These genes have a regulatory DNA sequence located in their upstream (at their 5'-end) region, which makes their expression dependent on light; this sequence is called ***light response element*** *(LRE).*

Some varieties of many of our crops, and indeed of many wild species, flower only when they receive either less than or more than a specified number of hours of light every 24 hours. Such varieties are termed as ***photosensitive.***

Nutrition

- *Biochemical mutants* : unable to synthesize a specific biochemical.
- Bacterial genes induced by availability of the concerned substrate, *e.g., lac* operon genes by lactose.
- *Stringent response* : drastic reduction in RNA, protein etc. syntheses under poor nutrition.
  - — Due to synthesis of pppGpp by RelA.
  - — Quickly reversed by adequate nutrition.

15.4.3. Nutrition

Proper growth and development of organisms depends on the availability of sufficient quantities of appropriate nutrition. The grain yield of a crop, milk yield of cattle and eggs laid by poultry are all greatly affected by the quality and the quantity of nutrition provided. Many mutants of microorganisms are unable to produce some biochemicals normally produced by their wild type strains. Such mutants show normal growth and development in the presence of the particular biochemical they are unable to synthesize, while they fail to grow if it is not supplied. Such mutants are known as ***biochemical*** or ***auxotrophic mutants***. Different biochemical mutants of *Neurospora* are unable to sunthesize any one or more of the following: an amino acid, a vitamin, adenine, uracil, urea etc. These mutant genes also show a threshold since their effect can be observed only in the absence of the specific biochemical from the medium (environment) in which the mutants are maintained.

In bacteria, expression of genes concerned with substrate utilization occurs only when the specific substrates are available to the cells. For example, genes for lactose utilization are expressed only when lactose is present in the medium. In such cases, the substrate molecules bind to specific regulator proteins and thereby, prevent their binding to specific operator DNA sequences: this allows the genes of concerned operons to be transcribed (Chapter 31). Similarly, when *E. coli* cells are provided a mixture of glucose and other sugars, they utilize exclusively glucose. This occurs because in the presence of glucose genes for the utilization of other sugars are not expressed; this is called ***catabolite repression*** (Chapter 31).

When bacterial cells encounter a poor growth condition, they show a drastic reduction in the syntheses of RNA, proteins, lipids etc.; this is called ***stringent response.*** This response is mediated by RelA, which produces an unusual nucleotide

pppGpp, which is converted into ppGpp by several enzymes. This nucleotide (ppGpp) interacts with several regulatory proteins and modifies their functions, which ultimately generates the stringent response. The stringent response is reversed rapidly on improvement in the growth conditions.

Humidity and Water
- Adequate water essential for normal growth and development.
- Water stress produces ABA, which induces specific genes

Maternal Relations
- Genotype of mother affects growth and development of embryos.
- Rh — incompatibility (haemolytic anaemia of babies

Parasites
- Produce disease; differentiate between susceptible and resistant individuals.
- *avr* gene of pathogen induces expression of PR proteins leading to HR response.

15.4.4. Humidity and Water

Optimum levels of soil moisture and humidity are essential for appropriate growth and development of organisms. The yield potential of high yielding varieties is fully expressed only when they are given optimum levels of irrigation; more or less irrigation water is detrimental to yield. Similarly, disease development in disease susceptible varieties depends on suitable temperature and moisture in the environment. Some varieties of cereals are lodging resistant, while others are susceptible to lodging. The susceptible varieties lodge only when appropriate conditions for lodging (high soil moisture and strong winds) are present. Similarly, drought tolerant genotypes can be identified from those susceptible to drought only under the conditions of limited water supply.

In plants, ABA (abscisic acid) is induced under conditions of water stress. A number of proteins are induced by (produced in response to) ABA; some of these proteins, *e.g.*, dehydrin, osmotin etc., confer protection from moisture stress. This is similar to the situation found in animals, where growth factors induce expression of specific proteins.

15.4.5. Oxygen

Expression of some genes is regulated by the presence/absence of oxygen. For example, alcohol dehydrogenase 1 (*Adh1*) gene of maize is expressed only under anaerobic conditions. Gene *Adh1* has an enhancer (actually, silencer), which suppresses its expression in the presence of oxygen. Under anaerobic conditions, *Adh1* is expressed simply because its silencer is unable to prevent its expression.

An ***enhancer*** is a DNA sequence, which markedly increases the expression of the concerned gene by affecting the activity of the promoter of that gene. A ***silencer***, on the other hand, suppresses the expression of the concerned gene. The enhancer/silencer sequences may be located on either side of the promoters they affect, and some of them may be separated from the promoters by several kb (klio base pairs) of DNA. All the genes expressed either in response to a stimulus or in specific tissues have appropriate enhancer/silencer sequences.

15.4.6. Maternal Relations

During their early development, most of the higher plants and animals are dependent on their mothers in one way or the other. Consequently, growth and development of embryos in these organisms is greatly affected by the genotypes of their mothers. For example, seed size, protein content and many other seed characteristics of crop plants are affected by the genotypes of plants producing the seeds. In human beings, if mother has Rh^- blood group and its foetus is of Rh^+ blood group, some cells from the foetus may pass into the blood stream of the mother, particularly at the time of child birth. This induces production of antibody-Rh° (Rh° = the antigen present on Rh^+ RBCs) in the mother, which leads to a special kind of haemolytic anaemia in the subsequent developing foetuses. This anaemia is not produced by the genotype of the foetus itself but it is the consequence of an interaction of the foetus's genotype with that of its mother.

15.4.7. Parasites

Organisms are infected by a variety of pathogens, *viz.*, fungi, bacteria, viruses and nematodes, and infested by insect pests. The ability of parasites to infect/infest the host is governed by their genotypes, while the host genotype determines its ability to resist the invasion by parasites. Thus different varieties of the host are classified as resistant or susceptible depending on their ability to resist the infection by a parasite. In order to identify susceptible varieties from resistant ones, it is necessary that both be exposed to the appropriate parasite. Thus the expression of disease resistance and susceptibility requires a specific biotic environment as a threshold.

In case of plants, many cases of disease resistance fall in the category of ***hypersensitive response*** (HR) in which a group of cells surrounding the site of infection by the pathogen dies along with the invading pathogen. HR is governed by a plant gene, *R*, for disease resistance. However, HR is produced only when the product of an *R* gene interacts specifically with its corresponding *a*vr (*avirulence*) gene product (or a compound produced by this gene product) of the invading pathogen.

Following this highly specific interaction, *R* gene product induces the expression of a set of genes encoding ***PR (pathogenesis-related)*** proteins, which give rise to the HR response. Thus expression of genes governing *PR* proteins and the consequent HR is dependent on a specific stimulus from the invading pathogen (Fig. 15.1).

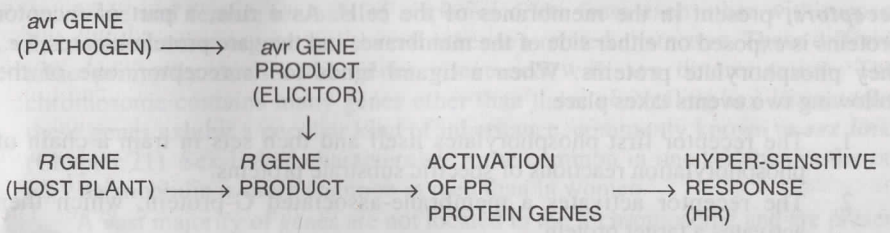


Fig. 15.1. A schematic representation of the expression of HR in response to the *a*vr gene product (or a compound produced by this product) of the pathogen; this protein/compound is called elicitor.

Symbionts
- Help in expression of host genes for symbiosis.

Population Density
- In many gram – negative bacteria; certain genes are expressed only at a threshold population density.
- Due to an autoincucer molecule produced by the concerned bacteria.

15.4.8. Symbionts

In ***symbiosis,*** both the host and the symbiont benefit from each other. An example of a familiar symbiosis is that between legumes and the bacterium *Rhizobium*. This bacterium forms nodules on legume roots, obtains its nutrition from the host and, in turn, provides the host with ammonium obtained by reducing atmospheric nitrogen. Both nodulation (formation of nodules) and nitrogen fixation (reduction of atmospheric nitrogen to ammonia) are affected by genes of the legume host as well as *Rhizobium*. Some mutant strains of the host are nonnodulating (*nod*$^-$), while others are defective in nitrogen fixation (*fix*$^-$). These strains can be identified from the normal nodulating and fixing strains of the host only when both of them are allowed to be nodulated by a specific *Rhizobium* strain.

15.4.9. Population Density

In many gram-negative symbiotic / pathogenic and some free-living bacteria, certain genes are expressed only when their population density reaches a threshold level. This phenomena is called ***quorum sensing***. These genes can be transcribed (*i.e.,* produce mRNA) only when a regulator protein binds to their promoters. The regulator protein can bind to the promoter only when it associates with a specific compound called *autoinducer*. The bacterial cells themselves produce the autoinducer, but it freely diffuses out of them. As the density of bacterial population increases, there is a consequent rise in the concentration of the autoinducer in the immediate environment of the bacterial cells. Ultimately, the autoinduccer concentration reaches the threshold level and the genes are expressed. The genes subject to quorum sensing include *lux* genes of many bacteria, genes encoding extracellular enzymes, exotoxins etc.

Organisms have elaborate response systems to external stimuli, including various molecules. These molecules, called ***ligands***, bind to specific proteins, called ***receptors,*** present in the membranes of the cells. As a rule, a part of receptor proteins is exposed on either side of the membrane, and they are protein kinases, *i.e.,* they phosphorylate proteins. When a ligand binds to its receptor, one of the following two events takes place.

1. The receptor first phosphorylates itself and then sets in train a chain of phosphorylation reactions of specific substrate proteins.
2. The receptor activates a membrane-associated G-protein, which then activates a target protein.

In either case, ultimately specific genes are expressed in response to the ligand binding to the receptor. This is called ***signal transduction***. In case of *Agrabacterium*, acetosyringone (produced by plant cells) binds to the receptor VirA.

VirA then phosphorylates itself and then phosphorylates protein VirG, which makes it active. Active VirG causes the expression of all *vir* region genes, leading ultimately to the transfer into plant cells of the *Agrobacterium* T- DNA (Chapter 37).

15.5. INTERNAL ENVIRONMENT

Internal environment consists of those factors, which are within the body of the concerned organisms, *e.g.,* age, sex, substrate and genetic background.

Age
- Certain genes are expressed only at a particular age, *e.g.,* baldness at 20 -30 yrs.

Sex
- Gene expression affected by sex of the individuals:
  - (*i*) *Sex-influenced trait* : Dominance of an allele affected by sex; sheep horned (*H*) dominant in males, recessive in females.
  - (*ii*) *Sex-limited trait* : Trait expressed in only one sex; *e.g.,* secondary sex characters, milk production in cattle etc.

15.5.1. Age

All the genes of any organism are not expressed at the same time. Most of the genes of an organism are expressed during a particular stage of development, while some genes are expressed throughout the life of the organism; the latter are called ***house-keeping genes.*** For example, the genes affecting milk production in cattle and those governing egg laying in poultry are expressed only when these organisms attain a certain age. Many genes in human beings are expressed only after a certain age. Genes controlling blood groups are expressed even before the child is borne. The recessive genes producing alkaptonuria and phenylketonuria are expressed soon after birth, while that causing rickets is expressed after the child is about one year old. Genetic baldness is expressed only after the age of 20-30 years, diabetes melitus is expressed between the age of 40-60 years, while the expression of Huntington's chorea may occur any time between the age of 15 and 70 years. Similarly in plants, flower, fruit and seed characters can be expressed only when the plants enter the reproductive phase.

15.5.2. Sex

Male and female animals of a species differ from each other with respect to their reproductive organs and several other sex-related characters. These differences arise primarily due to the genes concerned with sex determination. The *X* chromosome contains many genes other than those controlling sex characteristics; these genes exhibit a peculiar kind of inheritance, commonly known as ***sex linkage*** (Chapter 21). Sex-linked characters are more common in one sex than in the other, *e.g.,* haemophilia is more common in men than in women.

A vast majority of genes are not located in the *X* chromosome and are present in equal number of copies in both males as well as females. But the expression of some of these genes is markedly different in the two sexes. The characters produced by such genes are grouped into the following two classes: (1) sex-influenced characters and (2) sex-limited characters.

15.5.2.1. Sex-Influenced Characters

Ordinarily, the same allele of a gene is dominant in both the males and females of a species. Thus dominance relationship at a locus is not affected by sex. But *in case of some genes, dominance relationships of their alleles are affected by the sex of individuals in which they are present. As a consequence, one allele is dominant in males of the species, while the other is dominant in females. Characters produced by such genes are referred to as* ***sex-influenced characters.*** Some examples of sex-influenced characters are: the presence of horns in sheep; red and mahogany coat colour in cattle; crown baldness (*b*), harelip, gout and white forelock (all dominant over the normal trait in men), spina bifida and forked spinal cord (both dominant over the normal condition in women) in human beings.

The inheritance pattern of sex-influenced characters can be explained with the help of horned condition in sheep. Both males and females of the Dorset breed have horns, while those of Suffolk are hornless. In F_1 from Dorset × Suffolk all the males are horned, while all the females are hornless. In F_2, a ratio 3 horned : 1 hornless is obtained in the males, while females show the ratio 3 hornless : 1 horned.

These findings are summarised in Table 15.1. They can be explained by assuming that the *allele producing horns (H) is dominant over that producing hornless condition (h) in the case of males, but it (H) behaves as recessive to the latter (h) in females.* As a result, the heterozygous (*Hh*) males will produce horns, while such females will remain hornless. According to this assumption, Dorset has the genotype *HH* (both males and females horned), whereas Suffolk is of the *hh* genotype. In F_1 (*Hh*) from the cross Dorset × Suffolk, males (*Hh*) will have horns, while females (*Hh*) will be hornless (since *h* is dominant over *H* in females). In the F_2 generation, males would show the ratio 3 horned : 1 hornless, while in females 3 hornless : 1 horned ratio would be obtained (Table 15.1).

TABLE 15.1

Inheritance of Horned Condition in Sheep. Horned Condition is Dominant over Hornless in Males, While the Reverse is the Case in Females

			Phenotype	
Population	*Genotype*	*Frequency*	*Males*	*Females*
Dorset,	*HH*	All	Horned	Horned
Suffolk,	*hh*	All	Hornless	Hornless
F_1	*Hh*	All	Horned	Hornless
F_2	*HH*	1/4	Horned	Horned
	Hh	2/4	Horned	Hornless
	hh	1/4	Hornless	Hornless

Similarly, crown baldness is produced by a gene *b*, which is dominant in males but recessive in females to the allele *B* producing the normal phenotype. As a consequence, heterozygous (*Bb*) males develop baldness, while such females are normal in appearance. A female will develop crown baldness only when it is homozygous *bb*. Therefore, *sex-influenced characters are markedly more common in one sex than in the other.*

15.5.2.2. Sex-Limited Characters

Generally, genes express themselves in both the sexes. But some genes express themselves in only one of the two sexes although they are present in the individuals

of both the sexes in the appropriate genotype. Those characters whose development is limited to only one sex are known as ***sex-limited*** characters. Examples of sex-limited characters are the secondary sex characters in animals. These characters are confined to one sex only and serve as the basis for classification of individuals into males and females. For example, in *Drosophila,* maleness is produced by genes present in autosomes, while genes governing femaleness are located in the *X* chromosome. But males and females have autosomes as well as *X* chromosomes. Thus they both have genes for maleness as well as femaleness, but only one or the other is expressed making them either male or female.

Other examples of sex-limited characters are milk production in cattle, egg production in poultry, prolapsus of uterus in human beings etc. Genes governing milk production in cattle and those controlling egg production in poultry are contributed by both male and female parents of an individual. Obviously, these genes are present in both males and females, but they are expressed only in females simply because males of these species can neither produce milk nor lay eggs. Similarly, the gene for uterus prolapsus cannot be expressed in males since they do not have a uterus.

Substrate
- Several genes function in a sequence.
- The product generated by the activity of one enzyme serves as substrate for the other.
- Gives rise to various gene interactions.

Tissue -Specific Expression
- Genes expressed only in specific tissues.
- Due to enhancer sequences.

Genetic Background
- Modifying genes increase/decrease gene expression.
- Produces quantitative variation in oligogenic traits.
- Also affects dominance.

15.5.3. Substrate

In general, a gene expresses itself by producing a specific enzyme, which catalyses a specific biochemical reaction. However, an enzyme can catalyse the reaction only when the substrate for the reaction is available to it. These substrates are either obtained from the breakdown of the food consumed by the organism or are produced by enzymes encoded by other genes of the organism. The latter situation gives rise to gene interactions discussed earlier (Chapter 13).

The effect of availability of substrate on gene expression can be elaborated by the following example. In human beings, diabetes mellitus is the result of a deficiency of insulin, the hormone that regulates blood sugar level; some forms of the disease are hereditary. The blood sugar level of diabetics shows an abnormal increase after food, particularly carbohydrate and sugar, intake and remains abnormally high even during fasting. In advanced stages of the disease, varying amounts of sugar are secreted in urine. The levels of sugar in blood as well as in urine of patients can be regulated by regulating the quantity of fats and carbohydrates in their diets. Patients fed on a restricted fat and carbohydrate diet

show a much less intensity of the symptoms of this disease. But patients consuming large amounts of these foods show more intense symptoms of the disease.

15.5.4. Tissue-Specific Expression

A large number of genes are expressed only in specific tissues of plants/animals; these genes are called ***tissue-specific genes***. For example, seed storage protein genes are expressed in developing seeds and the gene *rbcS* (encoding the small subunit of enzyme RuBISCO) is expressed in leaves. The tissue-specific expression of these genes is primarily due to their promoter/enhancer sequences. For example, the LRE (light response element) sequences of *rbcS* gene also determines the leaf-specific expression of this gene.

15.5.5. Genetic Background

A gene does not produce a character by itself; it only produces the enzyme, which catalyses one of the reactions of a reaction chain that ultimately produces the character. It is, therefore, expected that the expression of almost every gene will be affected by the other genes present in the organism.

Expression of many oligogenes is affected by modifying genes. ***Modifying genes*** are a group of genes, which enhance or reduce the phenotypic expression of an oligogene; these genes have small, cumulative effect on the gene expression. However, they do not affect the concerned character on their own; they seem to only modify the expression of the concerned major gene. Modifying genes are known to affect the expression of a gene causing piebald spotting in mice, rats, guinea pigs, rabbits etc.; of self-incompatibility and male sterility genes in plants, genes for disease resistance etc. Thus modifying genes generate a continuous variation in a character governed by an oligogene.

Modifying genes are also known to affect the dominance relationship between alleles of a gene. For example, in moth *Abraxas*, allele *L* of gene *lutea* produces yellow colour and is incompletely dominant over allele *l*, which specifies white colour. Conscious selection was able to produce some moths in which *L* was completely dominant over *l* and some other moths in which *l* was completely dominant over *L*.

Phenocopy Effect

- Environment changes expression of one allele of a gene to such a degree that the phenotype due to another allele is produced.
- *Example* : Raising of *YY Drosophila* larvae on medium having silver salts produces yellow body colour.
- *No change in genotype;* only phenotype changed.

15.6. PHENOCOPY EFFECT

Certain specific environments change the expression of an allele of a gene to such an extent that the resulting phenotype is the same as that produced by another allele of the same gene. This type of effect due to the environment is referred to as ***phenocopy effect***, while the individuals possessing such altered phenotypes are called ***phenocopies***.

The phenocopy effect represents an alteration in the phenotype due to a modification of the expression of the concerned gene. It does not represent a change

in the gene itself, i.e., *gene mutation.* Consequently, *a phenocopy effect is usually confined to the individuals and the generation in which it is produced. It is generally not transmitted to the next generation unless the environment producing the phenocopy is also present in the next generation as well* (Fig. 15·2).

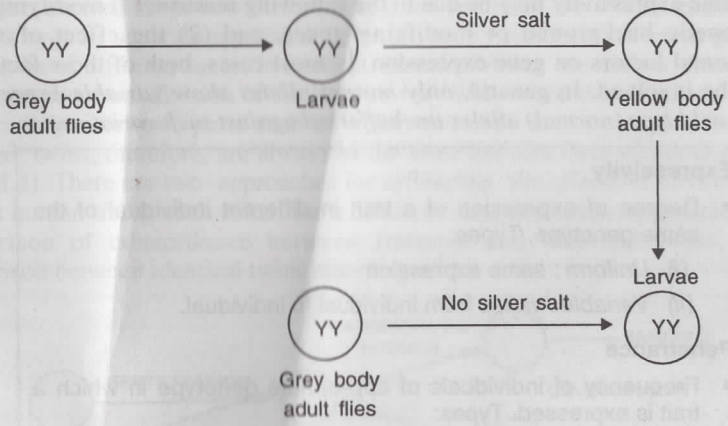


Fig. 15.2. Production of yellow body phenotype of *Drosophila* by the presence of silver salts in the medium on which larvae are raised.

In *Drosophila*, the recessive gene *y* produces yellow body colour, while its dominant allele *Y* produces the wild type gray body colour. When larvae with the genotype *YY* are raised on a medium containing silver salts, they develop into adult flies with yellow body colour. This alteration in body colour does not represent a genetic change. This can be verified by raising the next generation from such yellow body flies. When larvae from these yellow-body flies are raised on a normal medium devoid of silver salts, they develop into adult flies with gray body colour (Fig. 15.2). Clearly, the yellow body flies produced in this experiment had the same genotype as the normal gray flies (*YY*). But the expression of gene *Y* was altered by the silver salts to the extent that the phenotype generally produced by its recessive allele *y* was obtained. Several similar examples of phenocopy effect are known in both animals and plants.

15.7. EXPRESSIVITY

Most genes produce identical phenotypes in all the individuals in which they are present in the appropriate genotype. For example, all the seeds having the *w* allele, governing seed shape in pea, in the homozygous state (*ww*) have uniformly wrinkled shape. Similarly, those seeds that have either *WW* or *Ww* genotype are uniformly round. The ability of a gene to produce the same phenotype in all the individuals carrying it in the appropriate genotype is known as ***uniform expressivity***. As opposed to this, many genes have ***variable expressivity*** in that they produce variable phenotypes in the individuals that have them in the appropriate genotype.

An example of variable expressivity is furnished by the recessive gene producing partial chlorophyll deficiency in the cotyledonary leaves of lima beans. Cotyledonary leaves of seedlings homozygous for this gene show variable degrees of chlorophyll deficiency. In some seedlings chlorophyll is absent from the entire area of cotyledonary leaves, in most others a variable part of the cotyledonary leaves

is lacking the chlorophyll, while the leaves of other seedlings show chlorophyll deficiency only at the margins or at the tips. A single gene with variable expressivity produces a number of phenotypes often producing a quantitative type of variation in the concerned character.

Variable expressivity may be due to the following reasons: (1) modifying effect of the genetic background or modifying genes, and (2) the effect of various environmental factors on gene expression. In most cases, both of these factors are likely to be involved. In general, *only mutant alleles show variable expressivity, while the wild type (normal) alleles are bufferred against such variations.*

Expressivity

- Degree of expression of a trait in different individual of the same genotype. Types:
  - (*i*) *Uniform* : same expression
  - (*ii*) *Variable* : varies from individual to individual.

Penetrance

- Frequency of individuals of appropriate genotype in which a trait is expressed. Types:
  - (*i*) *Complete* : expressed in all.
  - (*ii*) *Incomplete* : expressed in a fraction of individuals.
- Both the phenomena due to environmental effects on character expression.

15.8. PENETRANCE

In general, genes express themselves in all the individuals in which they are present in the appropriate genotype; this is known as ***complete penetrance***. But many genes do not produce the concerned phenotype in all the individuals, which carry them in the appropriate genotype; such a situation is known as ***incomplete penetrance***. For example, the recessive gene producing partial chlorophyll deficiency in the cotyledonary leaves of lima beans shows incomplete penetrance as well; it expresses itself only in 10% of the individuals. When a gene is present in the appropriate genotype, the per cent of individuals in which it is able to express itself is a measure of its ***penetrance***. Thus the chlorophyll deficiency gene in lima beans has a *penetrance of 10%*. Almost all the genes showing incomplete penetrance exhibit variable expressivity as well. Thus incomplete penetrance is, in fact, an expression of variable expressivity in that some individuals show such a small expression of the gene that the trait produced is not detectable.

15.9. EFFECTS OF ENVIRONMENT ON THE DEVELOPMENT OF HUMAN CHARACTERS

A reliable investigation of the effects of environmental factors on the development of different characters of an organism requires that individuals having the same genotype be reared under different environments and observations on them be recorded at identical stages of development. This is achieved with relative ease in almost all the plants and many animals, but is ordinarily not possible in human beings. This is because every human being, except identical twins, has a different genotype from the rest and experimental populations of human beings can not be

created as is the case with other organisms. Therefore, a study of environmental effects on the development of human characters is based on *twins, i.e.,* the two babies born to a mother from a single pregnancy.

Twins are of two types: (1) fraternal twins and (2) identical twins. ***Fraternal twins*** are derived from two different zygotes resulting from fertilization of two different eggs by separate sperms. As a result, such twins have different genotypes, and may be of different sexes. In contrast, ***identical twins*** are derived from a single zygote; an abnormal division of the embryo very early in the development divides a single embryo into two parts each of which develops into a complete individual. Identical twins, therefore, are always of the same sex and have identical genotypes (Fig. 15.3). There are two approaches for estimating the effects of environment on various human traits, each suited for a different situation; these approaches are (1) comparison of concordance between fraternal and identical twins, and (2) comparison between identical twins reared apart.

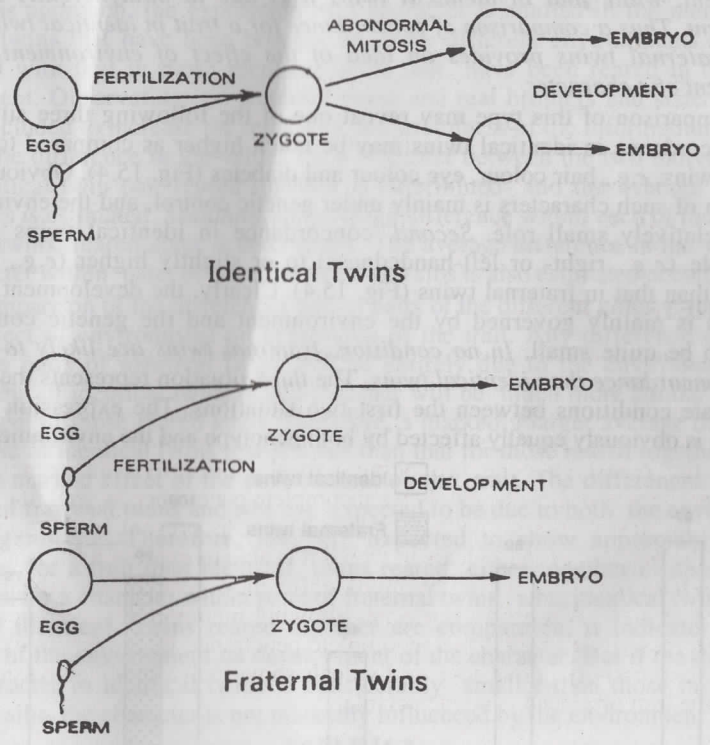


Fig. 15.3. Origin of identical and fraternal twins in man.

15.9.1. Concordance between Fraternal and Identical Twins

When a character is present in both the individuals of a twin pair, the pair is classified as ***concordant*** and the condition is known as ***concordance***. On the other hand, when the two individuals of a twin pair show different forms of the character, the pair is classified as ***discordant***, and the situation is referred to as ***discordance***. The numbers of twin pairs showing concordance and discordance may be recorded

and expressed as per cent of the total number of pairs studied; this is known as the ***frequency of concordance and discordance***. A comparison of the frequencies of concordance for a character in identical with that in fraternal twins provides some information on the effect of environment on the development of that character. It may be noted that in such studies, only those fraternal twins in which the two individuals in a twin pair are of the same sex are included.

The two individuals of each identical twin pair have identical genotypes. Therefore, *discordance for a character within an identical twin pair will be entirely due to the effect of environment on development of the concerned character.* However, concordance for a trait in identical twins will be due to both heredity and environment. In contrast, the two individuals of each fraternal twin are expected to have different genotypes. As a result, discordance in fraternal twins will be due to both heredity and the environment, but concordance will be mainly due to the environment. Thus *concordance for a trait in fraternal twins is due mainly to the environment, while that in identical twins it is due to both heredity and the environment. Thus a comparison of concordance for a trait in identical twins with that in fraternal twins provides an idea of the effect of environment on the development of a character.*

A comparison of this type may reveal one of the following three situations. *First,* concordance in identical twins may be much higher as compared to that in fraternal twins, *e.g.,* hair colour, eye colour and diabetes (Fig. 15.4). Obviously, the expression of such characters is mainly under genetic control, and the environment plays a relatively small role. *Second,* concordance in identical twins may be comparable (*e.g.,* right- or left-handedness) to or slightly higher (*e.g.,* alcohol drinking) than that in fraternal twins (Fig. 15.4). Clearly, the development of such characters is mainly governed by the environment and the genetic component appears to be quite small. *In no condition, fraternal twins are likely to show a greater concordance than identical twins.* The *third* situation represents the various intermediate conditions between the first two situations. The expression of such characters is obviously equally affected by both genotype and the environment.

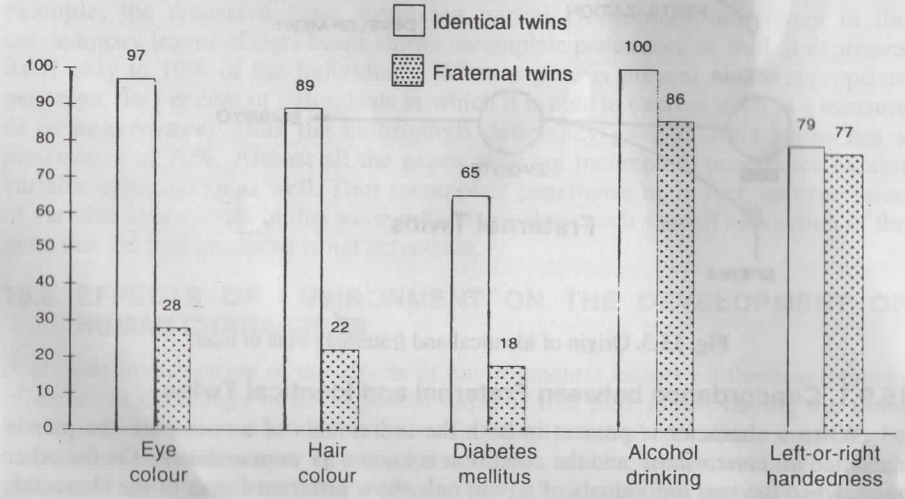


Fig. 15.4. Percentage of concordance for different characters in pairs of identical (unshaded bars) and fraternal (shaded bars) twins

This approach suffers from a number of limitations as follows. (1) No information is available on the mode of inheritance (oligogenic or polygenic) of the character. (2) The results are considerably affected by incomplete peneterance of genes. (3) Identical twins show a very high concordance for syndromes produced by chromosomal aberrations, although these syndromes are not produced by mutant genes. (4) A part of the concordance in fraternal twins is likely to be due to a similarity in their genotypes. (5) Identical twins are likely to receive a much more similar treatment than fraternal twins due to the almost total resemblance in appearances of the former. Thus, the environment may contribute more to the concordance in the case of identical twins than it does in fraternal twins. Finally, (6) this approach is applicable to qualitative characters only, which can be scored either as present or absent in the two members of a twin pair. As a result, it is not applicable to the study of quantitative characters.

15.9.2 Identical Twins Reared Apart

In this approach, the expression of a character is studied in the members of identical twins that have been brought up under different environments and it is compared with that in those identical twins that have been reared in the same environment. Observations on fraternal twins and real brothers and sisters (sibs) may be included to increase the reliability and usefulness of the information. In such a study, the difference in the values for a character between the two individuals of each identical twin pair reared together is determined and the average of these differences is estimated. Similarly, the average difference within each of the pairs of identical twins reared apart is determined. The average difference in the values of the character within pairs of fraternal twins and sibs is also estimated accordingly.

A comparison of the differences within pairs of the different kinds gives an idea of the effect of environment on expression of the trait. Any difference observed within the pairs of identical twins will be entirely due to the environment. The environment of identical twins reared together will be much more similar than the environment of those reared apart. Therefore, a markedly larger average difference in the case of identical twins reared apart than that for those reared together would indicate a marked effect of the environment on the trait. The differences between the pairs of fraternal twins and sibs are expected to be due to both the environment and the genotype. Therefore, they are expected to show appreciably larger differences for a trait than identical twins reared either together or apart. If the differences for a character within pairs of fraternal twins, sibs, identical twins reared apart and identical twins reared together are comparable, it indicates a large influence of the environment on development of the character. But if the difference for a character in identical twins is considerably smaller than those in fraternal twins and sibs, the character is not markedly influenced by the environment.

TABLE 15.2
Average Difference Between Individuals of Pairs of Identical Twins, Fraternal Twins and Sibs for some Quantitative Characters

Character	*Identical twins*		*Fraternal twins*	*Sibs*
	Reared together	*Reared apart*		
Height (cm)	1.7	1.8	4.4	4.5
Weight (lb)	4.1	9.9	10.0	10.4
Head length (mm)	2.9	2.2	6.2	
Head width (mm)	2.8	2.9	4.2	
IQ score (Binet)	5.9	8.2	9.9	9.8

The data listed in Table 15.2 may be explained as follows. (1) Identical twins reared together showed an average difference of 1.7 cm in their height, while those reared apart differed in height by 1.8 cm. Clearly, height is not markedly affected by the environment. This conclusion is supported by the fact that the average difference in the heights of fraternal twins was 4.4 cm, while that for sibs was 4.5 cm; these two values are comparable with each other but much larger than those for identical twins reared either together or apart. A similar situation is encountered for length and width of head. (2) In contrast, weight is greatly affected by the environment. Identical twins reared together and those reared apart exhibited the average difference of 4.1 and 9.9 lbs, respectively; these values indicate a marked effect of the environment on body weight. This conclusion is supported by the observations on fraternal twins and sibs, which showed a difference of 10.0 and 10.4 lbs, respectively; these values are comparable to those for identical twins reared apart. A similar picture is obtained for IQ scores as well.

15.9.3. Effect of the Environment on IQ Scores

IQ scores are used as measures of human intelligence. Although they suffer from several limitations, they have considerable practical usefulness. Several different IQ tests have been devised, but each suffers from one or the other deficiency. Numerous studies have been made on effect of the environment on IQ scores. The findings from these studies are summarised below. But it should be kept in mind that these findings do not reveal the mode of inheritance of IQ scores.

1. In general, children from smaller families show relatively higher IQ scores than do those from larger ones. This may be due to a greater attention provided by the parents in the case of smaller than in larger families. There is a considerable evidence that children receiving a greater attention from parents score higher in IQ tests than do those receiving less attention.
2. IQ scores of children from the same family are appreciably affected by their birth order. In general, a child borne earlier has a higher IQ score than does one borne later, the tests being conducted at a comparable age for both the children.
3. Persons from different regions, *e.g.*, north-south regions, of the same country show differences in their IQ scores. Similarly, individuals belonging to different human races (*e.g.*, Negroes, Mongoloids, etc.) and even from different religions show differences in IQ scores. These differences are mainly due to the differences in cultural values, economic conditions and the stress laid on education among these groups.
4. IQ scores are greatly affected by the guidance and the encouragement a person receives in his/her childhood. A child receiving proper guidance, prompt encouragement and suitable treatment would score higher in an IQ test than would one receiving poor guidance, no encouragement and illtreatment. Since children are more close to their mothers as compared to their fathers, the education, attitude and behaviour of mothers has a marked effect on the IQ scores of their children.
5. Malnutrition, especially of protein, during pregnancy and during early childhood is likely to adversely affect the IQ score of a child. Protein deficiency is known to be detrimental to growth and development of children.

QUESTIONS

1. What do you understand from environment ? Discuss the effects of various external environmental factors on gene expression.
2. Explain the meaning of internal environment. Discuss the effects of different factors of internal environment on the development of characters.
3. Explain the phenocopy effect with the help of a suitable example. How can you demonstrate that a phenocopy effect is not due to a gene mutation ?
4. Write short notes on the following:
   (*i*) threshold characters,
   (*ii*) expressivity,
   (*iii*) penetrance,
   (*iv*) sex-influenced characters,
   (*v*) sex-limited characters.
5. Persons living at high altitudes, say 4,800 m from sea level, have a considerably larger number (7.8 million/ml) of erythrocytes (red blood corpuscles) than those living at the sea level (4.25 million/ml). When person living at mean sea level move to a high altitude, the number of erythrocytes in their blood gradually increases to the level of those living at that altitude. Are the differences between erythrocyte counts in the bloods of persons living at different altitudes hereditary ?
6. In humans, stiff little finger (compactodactyly) is caused by a dominant gene with 75% penetrance. What proportion of offspring would show this trait if individuals heterozygous for this gene
   (*i*) marry normal individuals,
   (*ii*) marry heterozygous individuals,
   (*iii*) the penetrance was 50% in the above two cases,
   (*iv*) the penetrance was only 10% in the first two cases ?
7. Homozygous cock-feathered males are mated with homozygous hen-feathered females. In F_1, all the males are cock-feathered while, the females are hen-feathered. In F_2, the following results were obtained: males, 3 cock-feathered : 1 hen-feathered; females, 3 hen-feathered : 1 cock-feathered. How would you explain these findings ?

16

Inheritance of Quantitative Characters

16.1. INTRODUCTION

It was recognised quite early in the history of genetics that some characters showed discontinuous variation, while a majority of the biologically important traits exhibited continuous variation. In fact, most of the traits of economic importance in all the organisms show continuous variation. As a rule, characters showing continuous variation are subjected to measurements (of height, weight, time, value etc.). Therefore, data on such traits are referred to as ***measurement*** or ***quantitative data***, and such characters are termed as ***metric*** or ***quantitative characters.***

Quantitative Traits

- Continous variation.
- Studied by measuring their intensity.
- Markedly affected by environment.
- Genetic analysis :
  1. Not possible by classical Mendelian approach.
  2. Based on statistical procedures (*biometrical* or *quantitative genetics*).
- Governed by several genes with small additive effect.

The two characteristic features of quantitative characters are as follows: (1) continuous variation, and (2) a marked influence of the environment on their expression. Continuous variation makes their genetic analysis through the classical Mendelian approach virtually impossible. The large environmental effects on these traits make it very difficult to determine if the expression of a given character in an individual is due to heredity or environment.

16.2. HISTORICAL

The first serious attempt to study continuous variation was that by Francis Galton, a cousin of Charles Darwin. He used parent-offspring correlation and regression analyses to demonstrate that there was a predictable similarity between parents and their offspring with respect to quantitative characters. He presented this in a considerable detail in his book '*Natural Inheritance*' published in 1889.

In 1906, Yule theoretically demonstrated that if it was assumed that each of the several genes affecting a quantitative character had small and cumulative effect, and that they did not show dominance, they would produce a continuous variation. In 1903, Johanssen published the results from selection for seed size in rajma (*P. vulgaris*) and put forth the concepts of pureline, genotype and phenotype. These

results clearly showed that a part of variation in quantitative traits was heritable, but a substantial portion of it was due to the environment.

In 1908 and 1909, Nilsson-Ehle presented the multiple factor hypothesis to explain his results on the inheritance of seed colour in wheats and oats. Although seed colour does not show a truly continuous variation, it did demonstrate the existence of several factors, governing a single trait, with small cumulative effect and lacking dominance as postulated by Yule in 1906.

Subsequently, between 1910 and 1916, E.M. East and R.A. Emerson collected extensive data on true quantitative characters in tobacco and maize. These workers concluded that the inheritance of these traits closely followed the expectations from multiple factor hypothesis.

Multiple Factor Hypothesis

- Proposed by Nilsson - Ehle in 1908.
- Based on inheritance of grain colour in wheats and oats.
- Seed colour, in some crosses, governed by:
  1. Two genes, R_1 and R_2
  2. Alleles R_1 and R_2 produce colour (*positive alleles*).
  3. Alleles r_1 and r_2 do not produce colour (*negative alleles*).
  4. Each positive allele produces a small amount of colour.
  5. Effects of all positive alleles (of all the genes) are cumulative.
  6. Intensity of grain colour determined by the number of positive alleles of both R_1 and R_2.

16.3. MULTIPLE FACTOR HYPOTHESIS

The multiple factor hypothesis was postulated by Nilsson-Ehle in 1908. In studies on inheritance of seed colour in wheats and oats, he obtained 3 : 1, 15 : 1 and 63 : 1 ratios between coloured and white seeds from different crosses. It is clear from these ratios that the seed colour was governed by one (3 : 1 ratio in F_2), two (15 : 1 ratio in F_2) or three (63 : 1 ratio in F_2) genes. When two or three genes for colour were present together, they seemed to interact in the duplicate manner so that white seed colour was produced only when all the genes were present in the recessive state.

However, on a closer examination of the coloured seeds, Nilsson-Ehle found that there was a marked difference in the intensity of their colour. Therefore, he further classified the coloured seeds on the basis of intensity of their colour. The red seeds from crosses showing 15 : 1 ratio could be classified into four distinct classes: (*i*) dark red, (*ii*) medium-dark red, (*iii*) medium red and (*iv*) light red; they were present in the ratio 1 : 4 : 6 : 4. Thus what was apparently a 15 : 1 ratio turned out to be a 1 : 4 : 6 : 4 : 1 ratio on a closer examination. This was the first clear-cut demonstration for the existence of multiple factors, that is, genes with small and cumulative effects, and lacking dominance.

In order to explain the 1 : 4 : 6 : 4 : 1 ratio, Nilsson-Ehle made the following assumptions. (1) In crosses showing 15 : 1 ratio in F_2, seed colour is governed by two genes. (2) One of the alleles of each colour gene produces seed colour and is called ***positive allele***; it is generally denoted by a capital letter, *e.g.*, R_1, R_2 etc. The other allele of each colour gene does not produce any colour and is known as ***negative allele***; it is ordinarily denoted by the corresponding small letter, *e.g.*, r_1, r_2

etc. (3) These genes do not show dominance so that the heterozygote for a colour gene, *e.g.*, $R_1\ r_1$, is intermediate in colour between the two homozygotes ($R_1\ R_1$ and $r_1\ r_1$). This may be stated a little differently; two positive alleles of a colour gene ($R_1\ R_1$) produce twice the intensity of red colour of that produced by a single positive allele ($R_1\ r_1$). (4) Each of these positive alleles has a small, equal or almost equal, effect on seed colour. Finally, (5) the effects of positive alleles of the different colour genes are additive in the same manner as those of the two positive alleles of a single colour gene.

Thus *the intensity of seed colour depends on the number of positive alleles of all the genes affecting seed colour present in a seed* and not on which of these genes is present in the positive or the negative form. In other words, *a quantitative character is governed by several genes; the positive alleles of all the genes governing the trait are similar to each other in their action (effect), and that their effects are additive or cumulative in nature.* This is the essence of the multiple factor hypothesis.

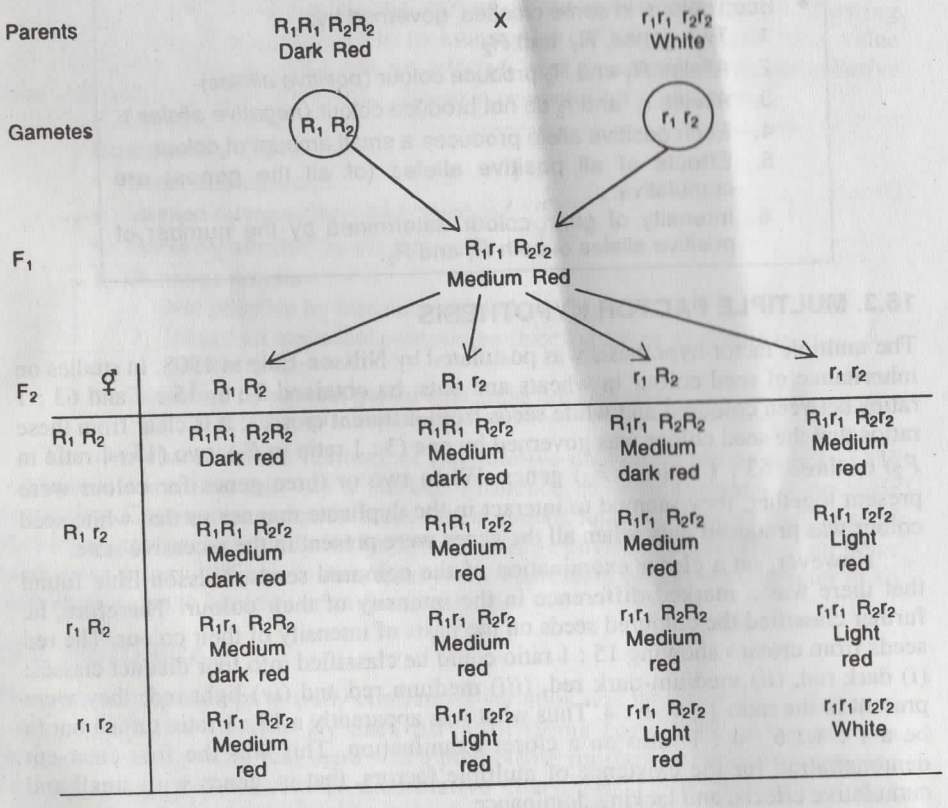


Fig. 16.1. Inheritance of seed colour in wheat

According to this hypothesis, inheritance of seed colour in crosses showing 1 : 4 : 6 : 4 : 1 ratio in F_2 may be explained as follows. The seed colour in such crosses is governed by two genes, say, R_1 and R_2 (positive alleles), r_1 and r_2 being the negative alleles of these genes. If the two parents of such a cross have the

phenotypes dark red and white, they would have the genotypes $R_1R_1\ R_2R_2$ and $r_1r_1\ r_2r_2$, respectively. These parents will produce the gametes R_1R_2 and r_1r_2, respectively, which will unite to give rise to the $R_1r_1\ R_2r_2$ genotype in F_1 (Fig. 16.1). Since F_1 seeds have only two positive alleles, their colour will be intermediate between those of the two parents, *i.e.*, medium red. In the F_2 generation of this cross, one zygote out of sixteen will have all the four positive alleles ($R_1R_1\ R_2R_2$); such seeds will be identical to one of the parents, and will have dark red colour. In four (out of 16) seeds, three positive alleles will be present ($R_1r_1\ R_2R_2$ and $R_1R_1\ R_2r_2$); these seeds will be lighter in colour than the coloured parent but darker than the F_1 seeds and will be medium-dark red in colour. Six (out of sixteen) seeds will have only two positive alleles each ($R_1r_1\ R_2r_2$, $R_1R_1\ r_2r_2$, $r_1r_1\ R_2R_2$); they will be identical with the F_1 seeds and medium red in colour. In four (out of 16) other seeds ($R_1r_1\ r_2r_2$ and $r_1r_1\ R_2\ r_2$), only one positive allele will be present; their colour will be light red, being intermediate between those of the F_1 seeds and the seeds of the white parent. In the remaining one seed (out of 16), there will be no positive allele ($r_1r_1\ r_2r_2$); this seed will be white in colour. Thus the multiple factor hypothesis is able to fully explain the 1 : 4 : 6 : 4 : 1 ratio for seed colour in the F_2 generation of wheat and oat crosses (Table 16.1).

TABLE 16.1

Genotypic and Phenotypic Frequencies Produced by the Segregation of Two Genes, R_1 and R_2 with Cumulative Effect on Seed Colour in Wheat

Genotype	*Frequency*	*Number of dominant alleles*	*Phenotype*	*Frequency*
$R_1R_1\ R_2R_2$	1	4	Dark-red	1
$R_1r_1\ R_2R_2$	2	3	Medium-dark red	4
$R_1R_1\ R_2r_2$	2	3	Medium-dark red	
$R_1r_1\ R_2r_2$	4	2	Medium red	6
$R_1R_1\ r_2r_2$	1	2	Medium red	
$r_1r_1\ R_2R_2$	1	2	Medium red	
$R_1r_1\ r_2r_1$	2	1	Light red	4
$r_1r_1\ R_2r_2$	2	1	Light red	
$r_1r_1\ r_2r_2$	1	0	White	1

Thus the studies of Nilsson-Ehle demonstrated the existence of several genes with small cumulative effects governing a single trait; such genes are commonly known as ***polygenes*** (sometimes, ***minor genes***). In contrast, genes having a large effect on the characters they govern are called ***oligogenes*** or ***major genes***. Oligogenes generally govern qualitative traits with discrete variation; that is, they permit classification of individuals into few discrete classes for the characters concerned.

However, variation in seed colour in wheats and oats was not continuous since the seeds could be classified into a few distinct groups on the basis of intensity of their colour. Characters governed by polygones (multiple factors), but showing discrete variation are called ***quasi-quantitative characters.***

Polygenes
- Govern quantitative traits.
- Each gene produces a small effect.
- Their effects are cumulative.
- Expression markedly affected by environment.
- Show additive, dominance and epistatic effects.

Oligogenes
- Govern qualitative traits.
- Each gene produces large effect.
- Produce discontinuous variation.
- Expression usually not much affected by the environment

16.4. STUDIES ON *NICOTIANA*

East and Emerson began, in 1910, independent evaluation of the multiple factor hypothesis to explain the inheritance of true quantitative traits in tobacco and maize, respectively. Statistical parameters like frequency distribution, mean, variance etc. were used to study the inheritance patterns of quantitative traits. Their studies clearly demonstrated that the inheritance pattern of quantitative traits followed the predictions based on multiple factor hypothesis. This clearly showed that even true quantitative traits were governed by multiple factors or polygenes.

In 1916, East reported his studies on the inheritance of corolla length in *Nicotiana longiflora*, which presented a conclusive evidence for polygenic inheritance of this trait. He crossed two inbred lines with contrasting corolla lengths of 40 and 93 mm. The F_1 from this cross was intermediate with mean corolla length of 63 mm. In F_2, a much larger variation than that present in the parents or the F_1 was observed for corolla length; this variation was continuous as well. The main features of his study are summarised below (Fig. 16.2).

1. The two parents were inbreds and almost homozygous. But there was variation for corolla length within each of them. Obviously, this variation in corolla length (within each parent) was due to the environment.
2. The F_1 mean was intermediate between those of the two parents. This is precisely what would be expected if corolla length were governed by multiple factors (polygenes) with cumulative effect.
3. The amount of variation in F_1 was comparable to that in the parents. Again all the F_1 plants will have the same genotype. Therefore, the variation for corolla length in F_1 also was produced by the environment.
4. The F_2 mean was close to that of F_1 and intermediate between those of the two parents. This is precisely what is expected in case of polygenic inheritance. In addition, there was a much greater variation for corolla length in F_2 than that in either of the two parents or in the F_1. Segregation and recombination of polygenes governing corolla length is expected to generate genetic variation in F_2. This together with the variation produced by the environment will account for the increased variation in F_2.

Thus it is clear that these observations correspond closely with the expectations based on Mendelian inheritance of polygenes. Confirmatory evidence for Mendelian

inheritance of corolla length may be obtained by growing individual plant progenies in F_3 and subsequent generations. Based on the Mendelian principles, the following predictions can be made.

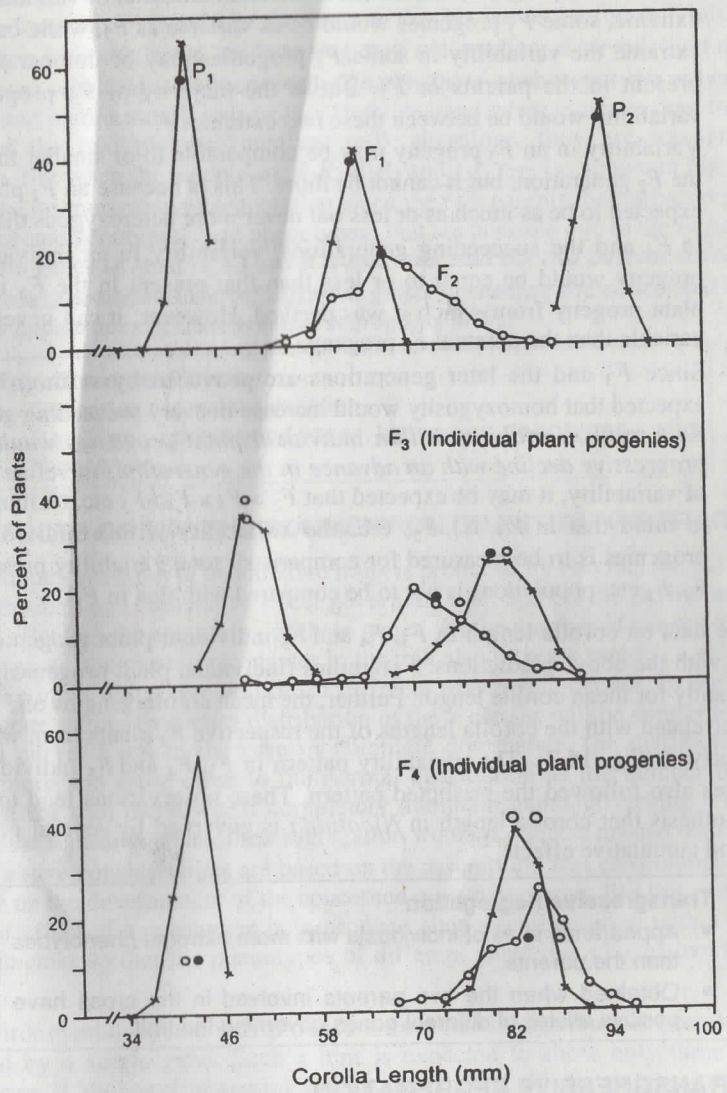


Fig. 16.2. Frequency distribution of the two parents, F_1, F_2 populations and F_3 and F_4 individual plant progenies for corolla length in *Nicotiana longiflora*. The closed circles represent the means of the various populations. In case of the F_3 and F_4 progenies, the open circle alongside the closed ones denote the corolla lengths of the parent plants of the previous generation from which the respective progenies were derived.

1. F_2 plants with markedly different corolla lengths should produce F_3 progenies with different mean corolla lengths, since such F_2 plants may be expected to have different genotypes for this trait.
2. Different F_3 progenies would show different amounts of variation. On one extreme, some F_3 progenies would be as variable as F_2, while on the other extreme the variability in some F_3 progenies may be comparable to that present in the parents or F_1. But in the majority of F_3 progenies, the variability would be between these two extremes.
3. Variability in an F_3 progeny may be comparable to or smaller than that in the F_2 generation, but it cannot be more. This is because an F_2 plant can be expected to be as much as or less but never more heterozygous than the F_1.
4. In F_4 and the succeeding generations, variability in an individual plant progeny would be equal to or less than that present in the F_3 individual plant progeny from which it was derived. However, it can never be more variable than the parental F_3 progeny.
5. Since F_3 and the later generations are produced by selfing, it may be expected that homozygosity would increase in every succeeding generation. As a result, *variability within individual plant progenies would show a progressive decline with an advance in the generation;* therefore, in terms of variability, it may be expected that $F_2 > F_3 > F_4 > F_5$ etc. (It must be kept in mind that in F_3, F_4, F_5, etc. the variability within individual plant progenies is to be measured for comparison; total variability present in F_3, F_4, F_5 etc. populations is not to be compared with that in F_2).

The data on corolla length in F_3, F_4 and F_5 individual plant progenies agreed closely with the above predictions. F_3 families (individual plant progenies) differed significantly for mean corolla length. Further, the mean corolla lengths of F_3 families were correlated with the corolla lengths of the respective F_2 plants from which they were derived (Fig. 16.2). The variability pattern in F_3, F_4 and F_5 individual plant progenies also followed the predicted pattern. These observations tend to confirm the hypothesis that corolla length in *Nicotiana* is governed by several genes with small and cumulative effects.

Transgressive Segregation
- Appearance in F_2 of individuals with more extreme phenotypes than the parents.
- Obtained when the two parents involved in the cross have positive alleles of different genes governing the trait.

16.5. TRANSGRESSIVE SEGREGATION

The appearance in F_2 of individuals with a higher or lower intensity of a character than those present in the parents involved in the cross is termed as ***transgressive segregation,*** and such individuals are called ***transgressive segregants.*** Such segregants are produced when (1) the two parents involved in a cross have positive alleles of different genes affecting a quantitative trait, and (2) segregation of these genes produces the two extreme homozygotes in F_2, which surpass the parental limits for the character.

For example, if the parents involved in the cross represented in Fig. 16.1 (for seed colour in wheat) had the genotype $R_1R_1\ r_2r_2$ and $r_1r_1\ R_2R_2$, they would have medium red seeds. The F_1 from these parents would have the genotype $R_1r_1\ R_2r_2$ and medium red seeds (which is the same as that depicted in Fig. 16.1). The F_2 generation from this F_1 would be the same as that shown in the figure. Thus in the F_2 generation from this cross, seeds having dark red, medium-dark red, light red and white colours would also be recovered. Clearly, these phenotypes are either more (dark red and medium-dark red) or less (light red and white) intense than the seed colours of the two parents (medium red); therefore, they are transgressive segregants. But when the parents are $R_1R_1\ R_2R_2$ and $r_1r_1\ r_2r_2$ (as shown in Fig. 16.1), there can be no transgressive segregation since the two parents themselves represent the two extreme genotypes (and phenotypes) that are possible due to the segregation of these two genes (R_1 and R_2). ***Thus it is essential that the two parents involved in a cross possess positive alleles of different genes governing the concerned trait to permit the appearance of transgressive segregants in*** $\mathbf{F_2}$.

Environmental Effects
- Generates continuous variation.
- Reduces correspondence between genotypes and phenotypes.

16.6. EFFECTS OF THE ENVIRONMENT ON QUANTITATIVE TRAITS

The continuous variation in quantitative traits is produced by two factors: (1) several genes with small and cumulative effect govern the trait, and (2) the environment has a marked effect on the expression of these genes. As the number of genes governing a trait increases, the phenotypic classes in the trait also increase rapidly, *e.g.*, with 12 genes affecting a trait, there will be 25 different phenotypic classes. In the case of additive gene action, frequency distribution of these classes will be similar to that of a normal curve. Even in the case of complete dominance, the distribution will become more and more close to the normal distribution as the number of genes governing the trait increases. If completely dominant genes governing a trait are linked in the repulsion phase, their segregation would produce a normal distribution.

The above considerations are based on the assumption that environment has no influence on the development of the concerned quantitative trait. But this is never so. The chief effect of environment is to produce phenotypic variation within a single genotypic class so that the phenotypes of different genotypic classes overlap each other.

Environmental effects will produce a continuous distribution even in a trait governed by a single gene. Such a trait is expected to show only three distinct phenotypes if the environmental effects were zero. If the expression of this monogenic trait is affected by the environment to the extent of 50%, the frequency distribution of this trait will be roughly similar to normal curve. If the contribution of environment increases to 75%, the distribution will be close to normal distribution. Experimental evidence suggests that in the case of most quantitative traits, the contribution of environment varies from 10-50%, while in some cases it may be much higher. Thus environmental effects would produce a continuous variation even if the number of genes governing a character was very small, or even one.

In real situations, *several genes govern a quantitative trait, and such traits are affected by the environment. The joint action of these two factors produces a continuous distribution in quantitative traits.*

16.7. QUANTITATIVE GENETICS

The continous variation present in quantitative or metric traits makes their genetic analysis through the classical approach (that is used for qualitative traits) impossible. Therefore, the inheritance of these characters is ordinarily studied using statistical procedures based on mean, variance etc. The concepts and the procedures of and the conclusions from studies on quantitative traits are dealt with by the branch of genetics called ***Quantitative*** or ***Biometrical Genetics***.

QUESTIONS

1. Briefly describe the multiple factor hypothesis of polygenic inheritance. How did the findings of Nilsson-Ehle demonstrate that quantitative traits are inherited according to the Mendelian laws ?
2. 'Environment plays an important role in the expression of quantitative traits'. Discuss this statement with the help of suitable examples.
3. Write short notes on the following:
   (*i*) multiple factor hypothesis,
   (*ii*) polygenes and oligogenes,
   (*iii*) effects of environment on quantitative traits,
   (*iv*) quantitative genetics.

17

Linkage and Crossing Over

17.1. INTRODUCTION

It is universally accepted that genes are located in chromosomes. During cell division, each chromosome behaves as a unit. It may, therefore, be expected that genes located in the same chromosome would move together to the same pole during cell division. As a consequence, such genes would fail to show independent segregation and would tend to be inherited together. This expectation was expressed by Walter Sutton in 1903 while propounding the chromosomal theory of inheritance.

A significant deviation from independent segregation was first reported by Bateson and Punnet in 1905 for flower colour and pollen shape in pea. They observed a 7 : 1 :1 : 7 ratio in a test cross in place of the normal 1 : 1 : 1 : 1 ratio expected from independent segregation. Bateson and Punnet proposed two new terms, coupling phase and repulsion phase. These terms are widely used even today. But they failed to explain the 7 : 1 : 1 : 7 ratio in the test cross.

Linkage : History
- Predicated by Sutton in 1903.
- Observed, but not understood by Bateson and Punnet in 1905.
- Correctly explained by Morgan in 1911.

The concept of linkage originated when Morgan demonstrated in 1910 that the *white eye (w)* gene of *Drosophila* is located in the *X* chromosome (***sex linkage***). Subsequently, he studied the inheritance of several other sex-linked genes. In 1911, he published the following conclusions from these studies. (1) Genes located in the same chromosome tend to stay together during inheritance; this tendency is called ***linkage***. (2) Genes are arranged in a linear fashion in the chromosomes. (3) The intensity of linkage between two genes is inversely related to the distance between them in the chromosome. He also proposed that (4) coupling and repulsion phases are two aspects of the same phenomenon, *i.e.*, linkage. Later studies have confirmed the above conclusions of Morgan.

17.2. LINKAGE

The tendency of two or more genes to stay together during inheritance is known as ***linkage.*** *Two or more genes are linked together because they are located in the same chromosome. Linked genes do not show independent segregation. As a consequence, the dihybrid ratios obtained in* F_2 *and test-cross generations are significantly different from the expected ratios of* 9 : 3 : 3 : 1 *and* 1 : 1 : 1 : 1, *respectively; this effect of linkage is more clearly noticeable in a test-cross*

generation. The frequencies of parental character combinations are markedly more than expected, while those of the new character combinations are considerably lower.

Linkage : Features

- Genes tend to stay together during inheritance.
- Significant deviation from 1 : 1 : 1 : 1 (testcross) and 9 : 3 : 3 : 1 (F_2) ratios.
- Genes arranged in a linear order in chromosomes.
- Linked genes located in the same chromosome.
- Linked genes recombine due to crossing over.
- Recombination frequency increases with the distance between linked genes.

17.2.1. Coupling Phase

In maize, a dominant gene *C* produces coloured seeds, while its recessive allele, *c*, determines colourless seeds. Another dominant gene *Sh* governs full seeds, whereas its recessive allele *sh* gives rise to shrunken seeds. When plants having coloured full seeds (*CC ShSh*) were crossed with those having colourless shrunken seeds (*cc shsh*), F_1 seeds were coloured full (*Cc Shsh*). The F_1 (*Cc Shsh*) was testcrossed with the double recessive strain (*cc shsh*). Out of the 8,368 seeds obtained from the test cross (*Cc Shsh* × *cc shsh*), 4,032 (48.2%) were coloured full, 4,035 (48.3%) were colourless shrunken, 149 (1.7%) were coloured shrunken and 152 (1.8%) were colourless full (Fig. 17.1).

Clearly, the four phenotypic classes are not present in the expected ratio of 1 : 1 : 1 : 1. The phenotypic classes coloured full and colourless shrunken have a much higher frequency than the expected 25%. These two character combinations are referred to as ***parental combinations, parental phenotypes*** or ***parental types*** since they are the same character combinations that were present in the two parents of the F_1. The remaining two phenotypic classes, coloured shrunken and colourless full, are far less frequent than the expected 25%. These two character combinations are called ***recombinant phenotypes*** or ***recombinant types*** since they are obtained by reshuffling of the characters present in the two parents of the F_1.

In the above example, it appears as if the two dominant genes *C* and *Sh* have a strong affinity for each other so that the frequencies of coloured full and colourless shrunken phenotypes are greater than expected. This situation is referred to as ***coupling phase.*** *It is due to the presence of genes* C *and* Sh *in the same chromosome..*

17.2.2. Repulsion Phase

Similarly, when plants having coloured shrunken seeds (*CC shsh*) were crossed with those having colourless full seeds (*cc ShSh*), the F_1 seeds were coloured full (*Cc Shsh*). But when the F_1 plants were test-crossed (*Cc Shsh* × *cc shsh*), 47.9% of the seeds so obtained were coloured shrunken, 49.1% were colourless full, 1.4% were coloured full and 1.5% were colourless shrunken.

In this case also, the parental types (colourless full and coloured shrunken) were more frequent, while the recombinant types (coloured full and colourless shrunken) were less frequent than expected (Fig. 17.2). It appears as if in this cross the

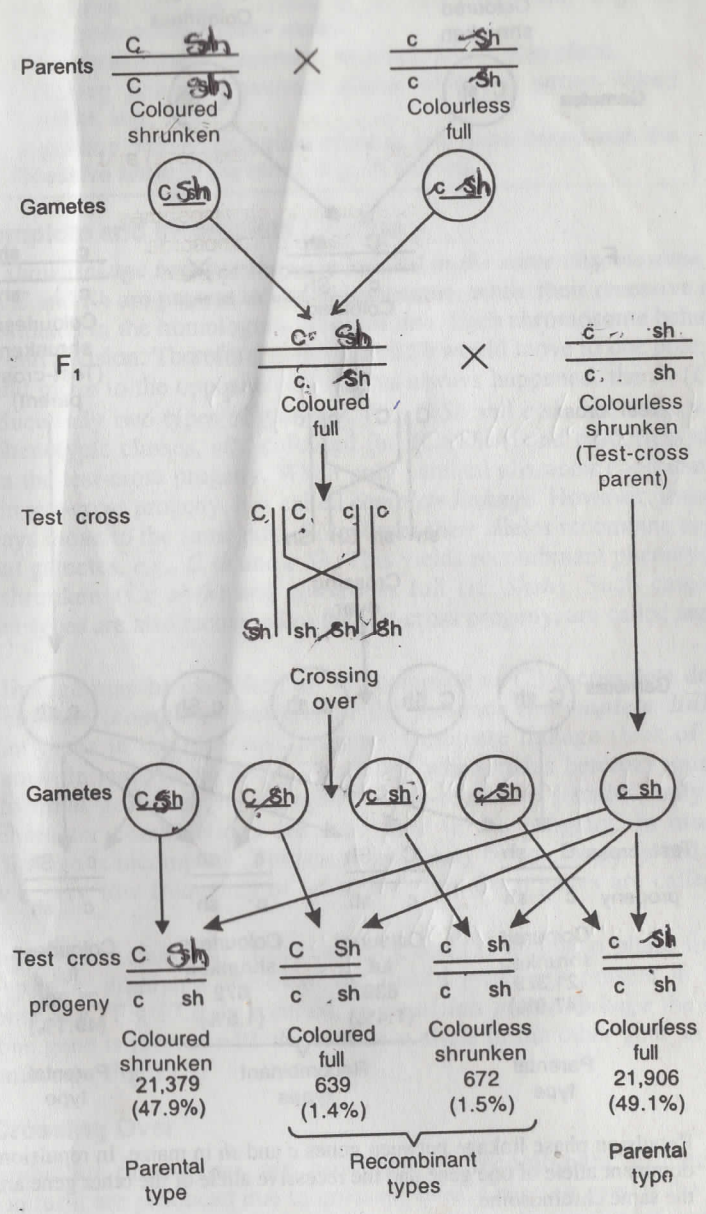


Fig. 17.1. Coupling phase linkage between genes *c* and *sh* in maize. In coupling phase, the dominant alleles of linked genes are located in one chromosome, while their recessive alleles are present in the homologous chromosome.

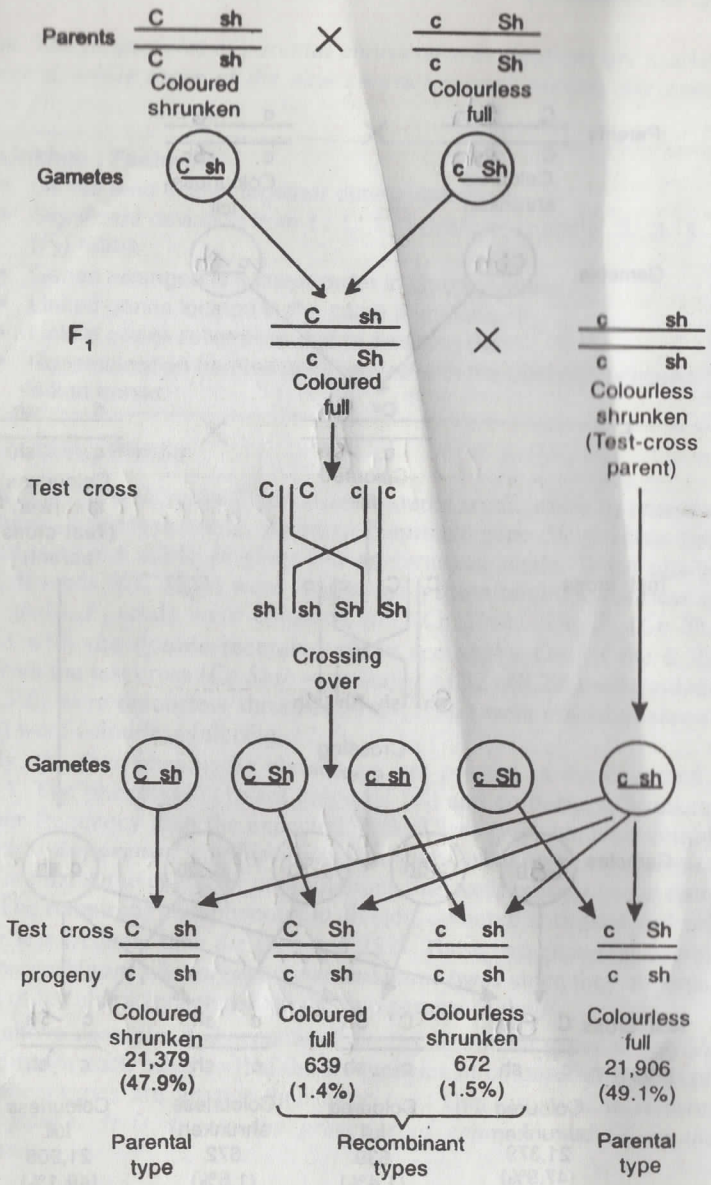


Fig. 17.2. Repulsion phase linkage between genes *c* and *sh* in maize. In repulsion phase, the dominant allele of one gene and the recessive allele of the other gene are present in the same chromosome.

dominant genes *C* and *Sh* dislike each other. This situation is referred to as ***repulsion phase.*** *It is due to the presence of the dominant allele of one gene, e.g.,* C, *with the recessive allele of the other gene, e.g.,* sh, *in the same chromosome. Obviously, coupling and repulsion phases are only the two situations of the same phenomenon of linkage.*

Linkage: Types

- *Complete linkage* : There is no recombination, *e.g.,* in *Drosophila melanogaster* males.
- *Incomplete linkage* : common; recombination takes place.
- *Coupling phase* : dominant alleles of linked genes linked together, *e.g., AB.*
- *Repulsion phase* : dominant allele of one gene linked with the recessive allele of the other, *e.g., A b* or *aB.*

17.2.3. Complete and Incomplete Linkage

Genes show linkage because they are located in the same chromosome. In Fig. 17.1, genes *C* and *Sh* are present in one chromosome, while their recessive alleles *c* and *sh* are situated in the homologous chromosome. Each chromosome behaves as a unit during cell division. Therefore, genes *C* and *Sh* would move to one pole, while *c* and *sh* would move to the opposite pole. If this always happened, the F_1 (*Cc Shsh*) would produce only two types of gametes, *viz.*, *C Sh* and *c sh*, and there would be only two phenotypic classes, *viz.*, coloured full (*Cc Shsh*) and colourless shrunken (*cc shsh*) in the test-cross progeny. When only parental character combinations are recovered in test-cross progeny, it is called ***complete linkage***. However, linked genes do not always move to the same pole. Sometimes their alleles recombine to produce recombinant gametes, *e.g., C sh* and *c Sh.* This yields recombinant phenotypes, *e.g.,* coloured shrunken (*Cc shsh*) and colourless full (*cc Shsh*). Such cases, where recombinant types are also recovered in the test-cross progeny, are called ***incomplete linkage.***

Thus linkage may be classified as (1) complete or (2) incomplete depending upon the absence (***complete linkage***) or the presence (***incomplete linkage***) of recombinant types in the test-cross progeny. Complete linkage (lack of crossing over) is known in male *Drosophila*. Therefore, when males heterozygous for two linked autosomal genes are mated with double-recessive females, only the two parental character combinations are recovered in the progeny. In most cases, however, linkage is incomplete. But some genes may be so closely linked that they may show a very low frequency of recombination. Such genes are called ***tightly linked.***

Linkage is also classified as (1) coupling and (2) repulsion phase linkage. In ***coupling phase,*** dominant alleles of the linked genes are present in the same chromosome, *e.g.,* Fig 17.1. In contrast, in ***repulsion phase linkage*** the dominant allele of one gene is present with the recessive allele of the other gene in the same chromosome, *e.g.,* Fig. 17.2.

17.2.4. Crossing Over

Recombinant phenotypes are produced by recombinant gametes. These gametes, in turn, are produced due to crossing over. ***Crossing over*** is the exchange of strictly homologous segments between nonsister chromatids of homologous chromosomes. Since genes are located in chromosomes, any recombination among linked genes must always be associated with the exchange of corresponding segments of the homologous chromosomes. This conclusion about crossing over is based on the genetic data. But it has been shown to be correct by some ingenious experiments.

17.3. SYMBOLS

Linkage between two genes is denoted by writing genotypes in one of the following four ways.

1. The genes present in one chromosome are written together. They are separated from those present in the homologous chromosome by an oblique line (/), *e.g.*, C *Sh* / *c sh* (Fig. 17.1) or *C sh* / *c Sh* (Fig. 17.2). *This system is easiest and the most convenient. It also occupies the least space. It is the universally adopted system.* (It may be noted that in case of independent assortment, these genotypes will be represented as *Cc Shsh* in both the cases.)
2. The genotype *C Sh* / *c sh* may be written as $\frac{C\ Sh}{c\ sh}$. Here the two lines represent the two chromosomes in which these genes are located.
3. The above genotype may also be written as $\frac{C\ Sh}{c\ sh}$. Here, only one line is drawn to represent the chromosome pair in which the genes are located.

   The systems 2 and 3 are quite cumbersome and occupy the greatest space. They are rarely followed.
4. The genotype *C Sh* / *c sh* may also be written as (*C sh*) (*c sh*). Here the genes present in one chromosome are placed within one parenthesis. This system takes greater space than system 1.

Linkage: Symbols and Detection

- *Symbol* : *AB*/*ab* (*A*, linked with *B*, and *a* with *b*), *Ab*/*aB*; most commonly used.
- *Detection* : Most easy in test cross:
  (*i*) The two parental types are in large excess of 25% each.
  (*ii*) The two recombinant types much smaller than 25% each.

17.4. DETECTION OF LINKAGE

Linkage between two dominant genes produces significant deviation from the typical dihybrid ratios of 1 : 1 : 1 : 1 (in test cross generation) and 9 : 3 : 3 : 1 (in F_2 generation). This deviation is the most easily detected in test cross data. In case of linkage, the two parental types are the most frequent; they are in excess of the expected 25% for each parental type. In contrast, the two recombinant types have markedly lower frequencies than 25% each. Thus, *it is typical of linkage that two phenotypic classes are markedly greater than 25% each, while the remaining two phenotypes are markedly lower than 25% each.* For example, in the test cross depicted in Fig. 17.1, the following data were obtained.

Coloured full	4,032	Parental types
Colourless shrunken	4,035	
Coloured shrunken	149	*Recombinant types*
Colourless full	152	

It may be seen that the two parental types have comparable frequencies. Similarly, the frequencies of the two recombinant types are also comparable. This type of test cross data is a sure indication of linkage.

As a rule, ***the parental types have much higher frequencies than the recombinant types*** . *In fact, this relationship can be used as a safe guide to identify the two types in the test cross data whenever the identity of the parental types is not known.*

17.5. CROSSING OVER

The exchange of homologous segments between nonsister chromatids of homologous chromosomes is known as ***crossing over.*** *It is responsible for recombination between linked genes. Crossing over takes place during pachytene.* In pachytene, each chromosome of a bivalent (chromosome pair) has two chromatids. Thus each bivalent contains four chromatids or strands (***four-strand stage***). Generally, one chromatid from each homologue is involved in crossing over (Fig. 17.3).

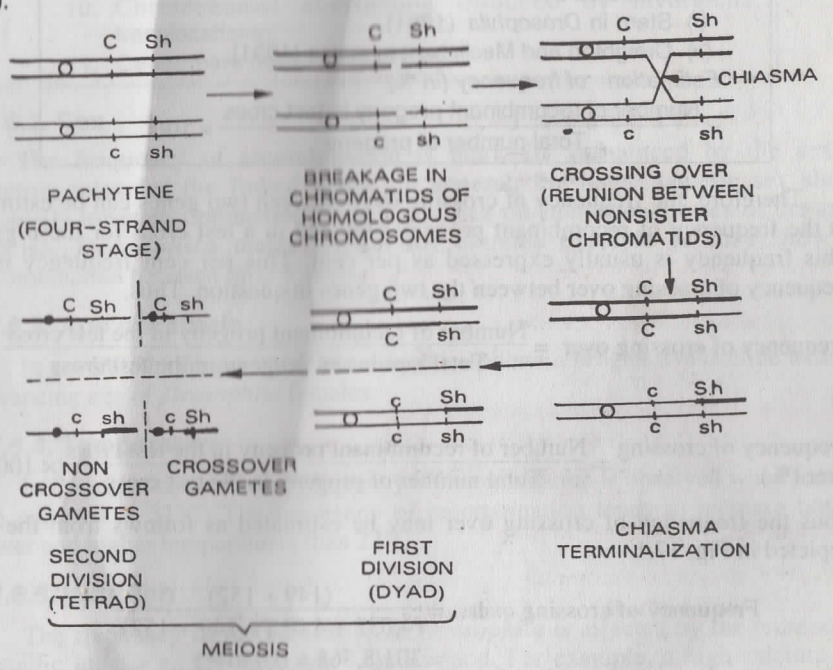


Fig. 17.3. Crossing over between nonsister chromatids of homologous chromosomes leading to recombination between linked genes. Generally, crossing over at a given point involves only two of the four chromatids of a bivalent. As a result, meiosis yields two crossover and two noncrossover gametes for one event of crossing over.

In this process, a segment of one chromatid becomes attached in place of the homologous segment of the nonsister chromatid and *vice-versa*. In Fig. 17.3, it is assumed that breaks occur at precisely homologous points in the two nonsister chromatids. This is followed by reunion of the acentric segments. This produces a cross (×)-like figure at the point of exchange of the chromatid segments. This figure is called ***chiasma.***

Obviously, each event of crossing over produces two recombinant chromatids (called ***crossover chromatids***) and two original chromatids (referred to as ***noncrossover chromatids***) (Fig. 17.3). The crossover chromatids will have new

combinations of the linked genes, *i.e.*, will be recombinant. Gametes carrying them will produce the recombinant phenotypes in test-cross. Therefore, these phenotypes are also called ***crossover types***. Similarly, the noncrossover chromatids will contain the parental gene combinations. The gametes carrying them will give rise to the ***parental phenotypes*** or ***noncrossover types.***

Recombination and Crossing Over

- *Recombination* : formation of new combination of linked genes; *e.g.*, *Ab* and *aB* from *AB* / *ab*.
- *Crossing over* : exchange of homologous segments between nonsister chromatids of homologous chromosomes.
- Recombination is the consequence of crossing over; demonstrated by:
  (*i*) Stern in *Drosophila* (1931).
  (*ii*) Creighton and Meclintock in maize (1931).
- *Estimation of frequency (in %)*

$$= \frac{\text{Number of recombinant progeny in test cross}}{\text{Total number of progeny}} \times 100$$

Therefore, the frequency of crossing over between two genes can be estimated as the frequency of recombinant progeny obtained in a test cross for these genes. This frequency is usually expressed as per cent. This per cent frequency is the frequency of crossing over between the two genes in question. Thus,

$$\text{Frequency of crossing over} = \frac{\text{Number of recombinant progeny in the test cross}}{\text{Total number of progeny in the test cross}}$$

or

$$\text{Frequency of crossing over (\%)} = \frac{\text{Number of recombinant progeny in the test cross}}{\text{Total number of progeny in the test cross}} \times 100$$

Thus the frequency of crossing over may be estimated as follows from the data depicted in Fig. 17.1.

$$\text{Frequency of crossing over} = \frac{(149 + 152)}{(4{,}032 + 4{,}035 + 149 + 152)}$$
$$= 301/8{,}368 = 0.036$$

Therefore, frequency of crossing over (%) = $0.036 \times 100 = 3.6\%$.

17.6. FACTORS AFFECTING RECOMBINATION FREQUENCY

The frequency of recombination between linked genes is affected by several factors. The effects of these factors are summarised below. This description is mainly based on data from *Drosophila*.

17.6.1. Distance between Genes

The frequency of crossing over between two genes is positively associated with the distance between their locations in the chromosome. Thus, crossing over between two genes would increase with an increase in the distance between them. Since the locations of any two genes in a chromosome are generally fixed or

constant, the recombination frequency between them may also be expected to show little variation.

Factors Affecting Recombination Frequency

1. Increases with *distance between genes.*
2. *Sex* (lower in heterogametic sex).
3. *Age* (declines with age).
4. *Temperature* (lowest at an optimal temperature).
5. *Nutrition* (increases by removal of metallic ions, starvation).
6. *Chemicals* (increased by antibiotics, alkylaking agents etc.).
7. *Radiations* (increased by *X*- rays, γ- rays).
8. *Plasmagenes* (suppress recombination).
9. *Genotype* (*e.g.*, genes causing asynapsis, desynapsis etc.)
10. Chromosomal aberrations (reduced by inversions, translocations).
11. Centromere (lower in its vicinity).

17.6.2. Sex

The frequency of recombination is markedly influenced by the sex of heterozygotes for the linked genes. In general, the heterogametic sex shows relatively lower recombination frequencies than the homogametic sex of the same species. *Drosophila* males present the extreme example as they show no recombination between linked genes.

17.6.3. Age of Female

In general, the frequency of recombination shows a progressive decline with the advancing age of *Drosophila* females.

17.6.4. Temperature

In *Drosophila,* the lowest frequency of recombination is observed when females are cultured at 22°C. The frequency of recombination tends to increase both at lower and higher temperatures than 22°C.

17.6.5. Nutrition

The frequency of recombination in *Drosophila* is affected by the presence of metallic ions, *e.g.*, Ca^{2+} and Mg^{2+}, in its food. For example, a high calcium diet reduces recombination. A removal of metallic ions causes an increase in recombination. An increase in recombination was also observed when larvae were starved during certain stages of development.

17.6.6. Chemicals

Treatment of *Drosophila* with certain chemicals leads to an increase in recombination. For example, injecting the females with certain antibiotics like mitomycin C and actinomycin D promotes recombination. Treatment of females with alkylating agents, such as, ethyl methane sulphonate (EMS) etc., also has a similar effect.

17.6.7. Radiation

An increased recombination is observed when *Drosophila* females and many

other organisms are irradiated with X-rays, γ-rays and some other radiations. Even *Drosophila* males show recombination following irradiation with X-rays.

17.6.8. Plasmagenes

In *Drosophila*, some plasmagenes (genes present in the cytoplasm) affect recombination frequency. Usually they reduce recombination. In bajra, the Tifton male sterile cytoplasm is also known to reduce recombination.

17.6.9. Genotype

Many genes are known to affect the occurrence as well as the rate of recombination. Some of these genes affect chromosome pairing and the formation of synaptonemal complex. Some others affect the events that take place after synapses. Genes for asynapsis and those for desynapsis are two such examples.

17.6.10. Chromosomal Aberrations

In *Drosophila*, paracentric inversions reduce recombination between the genes located within the inverted segment. Therefore such inversions are often denoted by *C* (crossover suppressor). Translocations also reduce recombination in the vicinity of breakpoints.

17.6.11. Distance from Centromere

Centromere tends to suppress recombination. Therefore, genes located in the vicinity of centromeres show a relatively lower recombination rate than those located away from them.

Crossing Over

- Occurs in 4-strand stage; *during pachytene.*
- One event involves only 2 of the 4 chromatids of a bivalent.
- Produces two crossover and 2 noncrossover chromatids.
- Leads to chiasma formation during diplotene.
- Involves breakage and reunion.
- *Molecular basis* : Hybrid DNA model based on one double-strand break.

17.7. CYTOLOGICAL BASIS OF CROSSING OVER

Since linked genes are located in the same chromosome, they would show recombination only when there is exchange of the concerned segments between the homologous chromosomes, i.e., ***crossing over***. Experimental evidence for this expectation was provided in 1931 independently by Stern in *Drosophila*, and by Creighton and McClintock in maize.

17.7.1. Experiments of Stern in *Drosophila*

Stern used a *Drosophila* female in which one *X* chromosome was shorter than normal. This chromosome had the recessive gene *car* (*carnation* eye colour) and the dominant gene *B* (*Bar* eye shape). The other *X* chromosome of this female was of normal length, but a segment of *Y* chromosome was translocated into its short arm. This chromosome had the dominant gene car^+ (wild type allele of *car;* produces dull red eye colour) and the recessive gene B^+ (wild type allele of *B*; produces normal ovate eye shape) (Fig.17.4). Stern test-crossed this female to a *car* B^+ male.

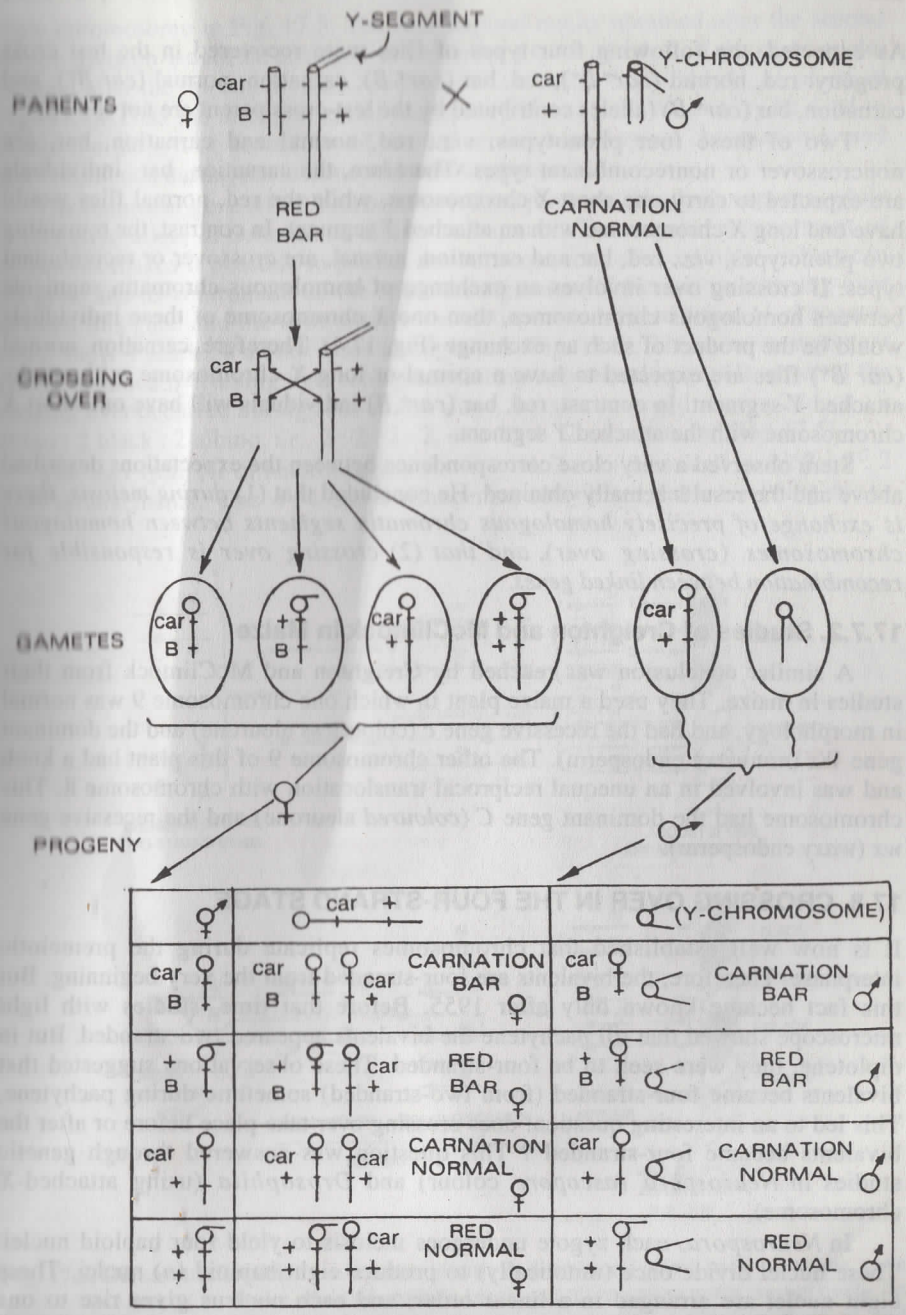


Fig. 17.4. Expected results from the experiment of Stern in *Drosophila* designed to establish the cytological basis of crossing over. These results are expected if recombination between linked genes involves the exchange of chromatin between homologous chromosomes. The results obtained by Stern showed a very high degree of agreement with those expected.

As expected, the following four types of flies were recovered in the test cross progeny: red, normal ($car^+ B^+$); red, bar ($car^+ B$); carnation, normal ($car\ B^+$); and carnation, bar ($car^+ B$) (alleles contributed by the test-cross parent are not shown).

Two of these four phenotypes, *viz.*, red, normal and carnation, bar, are noncrossover or nonrecombinant types. Therefore, the carnation, bar individuals are expected to carry one short *X* chromosome, while the red, normal flies would have one long *X* chromosome with an attached *Y* segment. In contrast, the remaining two phenotypes, *viz.*, red, bar and carnation, normal, are crossover or recombinant types. If crossing over involves an exchange of homologous chromatin segments between homologous chromosomes, then one *X* chromosome of these individuals would be the product of such an exchange (Fig. 17.4). Therefore, carnation, normal ($car\ B^+$) flies are expected to have a normal or long *X* chromosome without the attached *Y* segment. In contrast, red, bar ($car^+ B$) individuals will have one short *X* chromosome with the attached *Y* segment.

Stern observed a very close correspondence between the expectations described above and the results actually obtained. He concluded that (1) *during meiosis, there is exchange of precisely homologous chromatin segments between homologous chromosomes (crossing over), and that* (2) *crossing over is responsible for recombination between linked genes.*

17.7.2. Studies of Creighton and McClintock in Maize

A similar conclusion was reached by Creighton and McClintock from their studies in maize. They used a maize plant in which one chromosome 9 was normal in morphology, and had the recessive gene *c* (colourless aleurone) and the dominant gene *Wx* (*nonwaxy* endosperm). The other chromosome 9 of this plant had a knob and was involved in an unequal reciprocal translocation with chromosome 8. This chromosome had the dominant gene *C* (*coloured* aleurone) and the recessive gene *wx* (*waxy* endosperm).

17.8. CROSSING OVER IN THE FOUR-STRAND STAGE

It is now well established that chromosomes replicate during the premeiotic interphase. Therefore, the bivalents are four-stranded from the very beginning. But this fact became known only after 1955. Before that time, studies with light microscope showed that till pachytene the bivalents appeared two-stranded. But in diplotene, they were seen to be four-stranded. These observations suggested that bivalents became four-stranded (from two-stranded) sometime during pachytene. This led to an interesting question: does crossing over take place before or after the bivalents became four-stranded ? This question was answered through genetic studies in *Neurospora (ascopore* colour) and *Drosophila* (using attached-*X* chromosome).

In *Neurospora,* each zygote undergoes meiosis to yield four haploid nuclei. These nuclei divide once (mitotically) to produce eight haploid (*n*) nuclei. These eight nuclei are arranged in a linear order, and each nucleus gives rise to one ascospore. The eight ascospores produced form a zygote are enclosed within a membrance; the entire structure is called ***ascus***.

The arrangement of nuclei during meiotic and mitotic divisions and that of the mature ascospores is highly regular. As a result, the first meiotic division in the zygote *Al al* would produce two nuclei *Al* and *al*. The nuclei are represented by a

single chromosome in Fig. 17.5. The four haploid nuclei obtained after the second meiotic division would be arranged in the order *Al, Al, al, al*. And the eight nuclei obtained after the mitotic division would lie in the order *Al, Al, Al, Al, al, al, al, al* (Fig. 17.5). Thus the arrangement of ascospores in a mature ascus may be used to deduce the pattern of separation of the concerned homologous chromosomes and their chromatids during the meiotic and mitotic divisions.

The dominant gene *Al* of *Neurospora crassa* produces black ascospores, while its recessive allele *al* determines albino ascospores. The regular arrangement of ascospores makes it possible to predict the consequences of crossing over between this gene and its centromere on the spore arrangement. If there is no crossing over between this gene and the centromere, the ascospores obtained from *Al al* zygotes would show 4 black : 4 albino spore arrangement (Fig. 17.5). The same arrangement will be obtained if crossing over took place during two-strand stage. However, if the crossing over occurred during four-strand stage, the spores would show a 2 black : 2 albino : 2 black : 2 albino, *i.e.*, 2 : 2 : 2 : 2, or a 2 : 4 : 2 arrangement (Fig. 17.5).

The asci from *Al al* heterozygotes show 4 : 4, 2 : 4 : 2 as well as 2 : 2 : 2 : 2 spore arrangements. The 4 : 4 spore arrangement may be explained as being due to

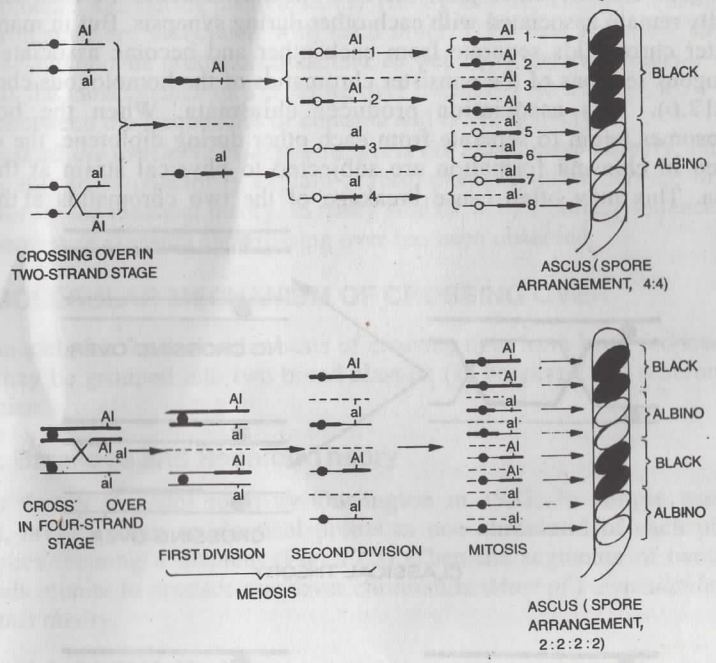


Fig. 17.5. Demonstration that crossing over occurs in four-strand stage. In *Neurospora*, gene *Al* and its recessive allele *al* produce black and albino ascospores, respectively. In the heterozygote, *Al al*, if there was no crossing over between gene *al* and centromere of the chromosome, or if crossing over occurred in two-strand stage, asci would show 4 black : 4 albino ascospore arrangement. But if crossing over occurred in four-strand stage, an arrangement different from 4 : 4 (*e.g.*, 2 : 2 : 2 : 2 or 2 : 4 : 2) would be obtained.

either no crossing over or crossing over in two-strand stage. But the 2 : 4 : 2 and 2 : 2 : 2 : 2 arrangements can be explained only on the basis of crossing over in the four-strand stage. Thus it may be concluded that crossing over occurs in the four-strand stage. ***It is universally accepted that crossing over takes place during pachytene when bivalents are in the four-strand stage.***

17.9. RELATIONSHIP BETWEEN CHIASMA AND CROSSING OVER

Chiasma is the cross-like (×) configuration observed during diplotene when homologous chromosomes begin to move away from each other. The homologous chromosomes appear to be attached with each other through chiasmata. Two divergent theories, *viz.*, (1) classical theory and (2) chiasma-type theory, have been advanced to explain the relationship between chiasma and crossing over.

17.9.1. Classical Theory

This theory, also called ***two-plane theory***, was advanced by Sharp in 1934. *According to this theory, a chiasma is formed when a chromatid of a chromosome becomes associated with the chromatid of its homologous chromosome.* When homologous chromosomes pair, the two sister chromatids of each chromosome generally remain associated with each other during synapsis. But in many sections, the sister chromatids separate from each other and become associated with the homologous sections of the nonsister chromatids of the homologous chromosomes (Fig. 17.6). This association produces chiasmata. When the homologous chromosomes begin to separate from each other during diplotene, the chromatids involved in chiasma formation are subjected to physical strain at the point of chiasma. This may often cause breakage of the two chromatids at this point. A

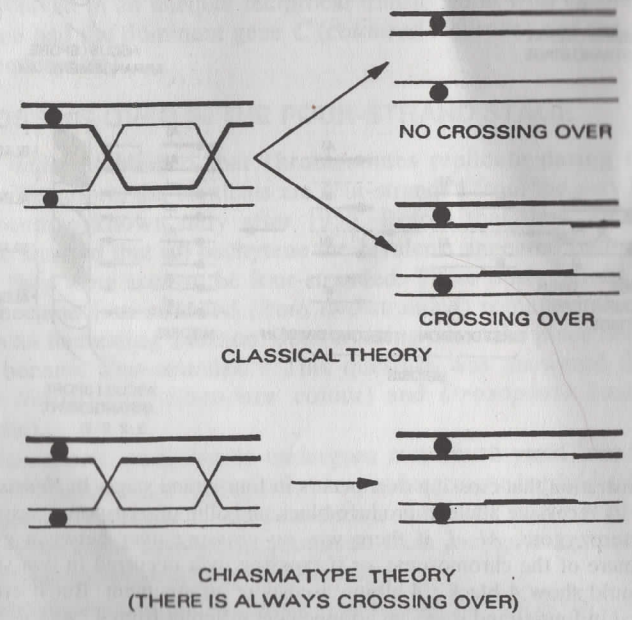


Fig. 17.6. The classical and chiasmatype theories of origin of chiasma and their consequences.

reunion between the chromatid segments thus produced would lead to crossing over or recombination between linked genes.

Thus according to classical theory, (1) a chiasma is formed when nonsister chromatids of homologous chromosomes become associated during synapsis, (2) chaisma formation is the cause of crossing over, (3) each chiasma does not lead to crossing over, and (4) crossing over occurs during diplotene. ***The available experimental evidence does not support this hypothesis and it is of only a historical significance.***

17.9.2. Chiasmatype Theory

This is also referred to as ***one plane theory***. It was first proposed by Janssens in 1909 and further expanded by him in 1924. Later, it was more fully developed by Belling and Darlington. According to this theory, breakage in nonsister chromatids of homologous chromosomes is followed by reunion of the chromatid segments so produced; this results in crossing over. When the homologous chromosomes begin to move away from each other during diplotene, chiasmata are formed at those points where crossing over had taken place earlier (during pachytene) (Fig. 17.6).

Thus, (1) chiasmata are produced due to crossing over. (2) Crossing over occurs before diplotene, *i.e.,* during pachytene. (3) Throughout the length of a bivalent, only sister chromatids are associated with each other. (Whereas nonsister chromatids become associated to produce chiasmata as per the classical theory). (4) Each chiasma is the consequence of a crossing over event. (5) As a result, a 1 : 1 ratio is expected between the frequencies of chiasma and crossing over.

Almost all the available evidence supports the chiasmatype theory. Further this theory is able to satisfactorily explain several phenomena, which are difficult to account for by the classical theory. In many studies, a 1 : 1 correspondence between the frequencies of chiasma and crossing over has been observed.

17.10. MOLECULAR MECHANISM OF CROSSING OVER

Several models for the molecular basis of crossing over have been proposed. These models may be grouped into two broad classes: (1) breakage and reunion, and (2) copy-choice.

17.10.1. Breakage and Reunion Theory

This theory was put forth by Darlington in 1932. In simple words, after synapsis, breaks occur at identical points in one chromatid of each of the two homologues forming a bivalent (Fig. 17.3). Then the segments of two nonsister chromatids reunite to produce crossover chromatids. *Most of the available evidence favours this theory.*

17.10.2. Hybrid DNA Models

The breakage and reunion among nonsister chromatids must essentially be based on breakage and reunion of DNA double helices. Several models of breakage and reunion of DNA molecules have been proposed. These models are also called ***hybrid DNA models.*** The various hybrid DNA models are modifications of either the Whitehouse model proposed in 1963, or the Holiday model advanced in 1964. Of these, the Holiday model is relatively simpler and more attractive.

17.10.2.1. Models Involving Single Strand Breaks

According to ***Holiday model,*** an endonuclease produces single-strand nicks (cuts) at identical points in the two homologous DNA molecules (Fig. 17.7). The two strands of each DNA molecule separate from each other upto some distance from the point of 'nick'. The free strands now pair with the intact strand of the homologous DNA molecule. This produces a 'hybrid DNA molecule'. The two nicks present in these molecules are sealed by DNA ligase. The hybrid DNA undergoes reorientation to form a ×-shaped figure. One end of this × now rotates by 180°. Such hybrid DNA molecules have been actually isolated from bacteria and photographed with the help of electron microscope. An endonuclease now induces nicks in the two intact (not cut earlier) strands of the hybrid DNA molecule. This yields two recombinant DNA molecules. Each recombinant molecule has a nick, which is finally sealed by DNA ligase. *Most of the available evidence was once considered to be in favour of this model, and almost all of the enzymes and proteins involved in the process are known to occur in bacteria.* But it is now widely held that recombination involves double strand breaks.

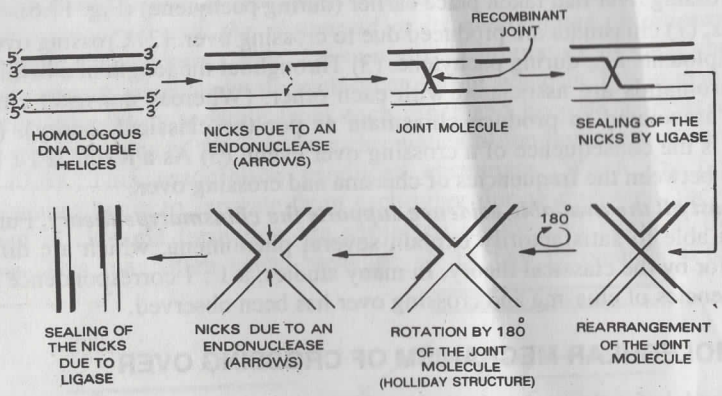


Fig. 17.7. Recombination due to single-strand breaks and reunion producing a hybrid DNA molecule; mainly based on Holiday model.

17.10.2.2. Model Based on Double Strand Break in One DNA Duplex

According to this model, a double strand break is induced by an endonuclease in one DNA duplex to initiate recombination; this duplex is called ***recipient-duplex.*** Now an exonuclease nibbles away one strand (from the 5'-end) of the duplex on either side of the break. This generates 3′- single-stranded ends on both sides of the break and the break is also enlarged into a gap. One of the single-stranded 3'-ends now invades the homologus region of the second DNA duplex, which is called the ***donor duplex.*** The homologous strand in the donor duplex is replaced by the invading strand; this generates a D loop in the donor duplex (Fig. 17.8).

The 3′- end of the invading strand is used as primer for repair synthesis of DNA using its complementary strand of the donor duplex as template; this enlarges the D loop. When the D loop becomes large enough, the displaced strand goes past the 3'- end at the other side of gap in the recipient duplex and pairs with the single - stranded region of this duplex. At this stage, the donor duplex has a heteroduplex region, while the recipient duplex has a single-stranded region (Fig. 17.8). This single-stranded region is filled by repair synthesis using the 3' - end as the primer.

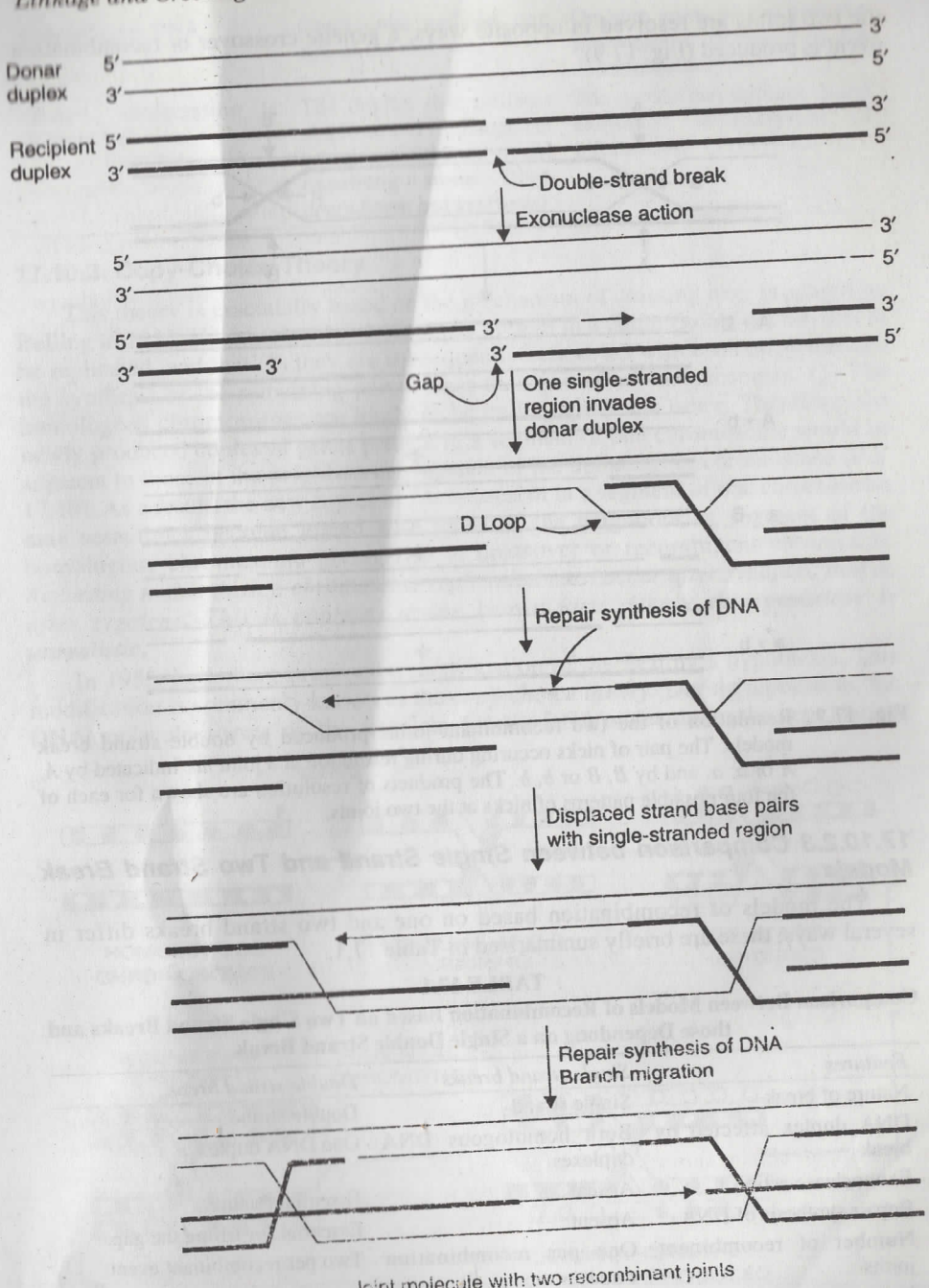


Fig. 17.8. Recombination model based on a double-stranded break in one DNA duplex.

The nicks remaining in the two duplexes are sealed by DNA ligase. The two DNA duplexes are now held together by two ×-like structures, called ***joints.*** When

the two joints are resolved in opposite ways, a genetic crossover or recombination event is produced (Fig. 17·9).

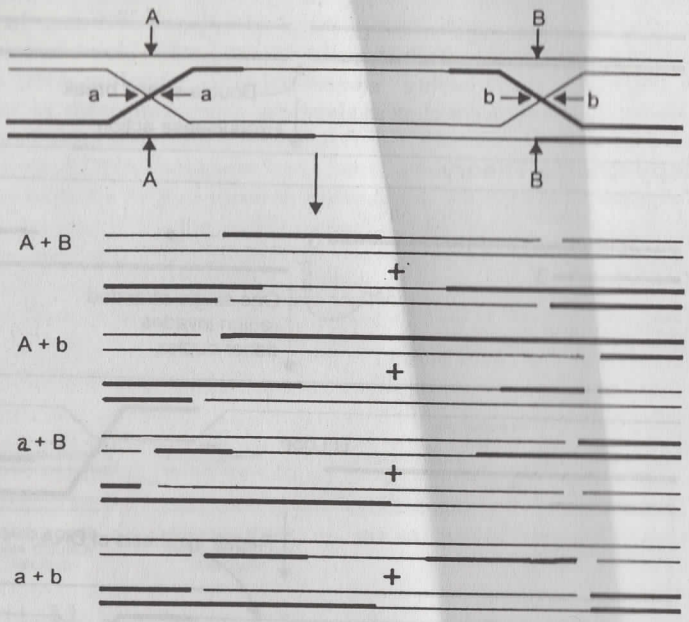


Fig. 17.9. Resolution of the two recombinant joints produced by double strand break models. The pair of nicks occuring during resolution at a joint are indicated by *A*, *A* or *a*, *a*, and by *B*, *B* or *b*, *b*. The products of resolution are shown for each of the four possible patterns of nicks at the two joints.

17.10.2.3 Comparison between Single Strand and Two Strand Break Models

The models of recombination based on one and two strand breaks differ in several ways; these are briefly summarised in Table 17.1.

TABLE 17.1
Comparison Between Models of Recombination Based on Two Single Strand Breaks and those Dependong on a Single Double Strand Break

Features	*Single strand breaks*	*Double strand break*
Nature of break	Single strand	Double strand.
DNA duplex affected by break	Both homologous DNA duplexes	One DNA duplex.
Exonuclease action	Absent	Essential feature.
Repair synthesis of DNA	Absent	Essential for filling the gap.
Number of recombinant joints	One per recombination event	Two per recombinant event.
Noncrossover products have	Heteroduplex regions in both duplexes	Heteroduplex regions in both the duplexes; the recipient duplex also contains a sequence from the donor duplex.
Heteroduplex DNA	Continuous and on one side of the nick site.	On both sides of the gap; the gap region is occupied by the donor sequence.
Genetic information is donated by	The duplex that initiates single strand exchange (in variants of the model requiring some DNA degradation and synthesis)	The duplex that initiates strand exchange is recipient of information.

17.10.3. Copy-Choice Theory

This theory is essentially based on the mechanism of crossing over proposed by Belling in 1933. He postulated that (1) genes present in a chromosome are the first to be replicated, and that (2) they are subsequently connected with each other through the synthesis of the remaining parts (other than genes) of chromosomes. (3) The homologous chromosomes are likely to be coiled with each other. Therefore, the newly produced copies of genes present in a segment of one chromosome would be adjacent to those of the neighbouring segment of the homologous chromosome (Fig. 17.10). As a result, the new copies of genes present in a segment of one chromosome may sometimes become joined with those of the neighbouring segment of the homologous chromosome giving rise to crossover or recombinant chromatids. *According to this theory, chromosome replication must occur after synapsis, that is, after zygotene. This is contrary to the known facts. Hence* ***this postulate is unrealistic.***

In 1955, Lederberg proposed a modification of the Belling's hypothesis. This modification is commonly known as the ***copy-choice theory.*** During replication, the DNA molecule serves as the template for the DNA molecule (or the copy) being

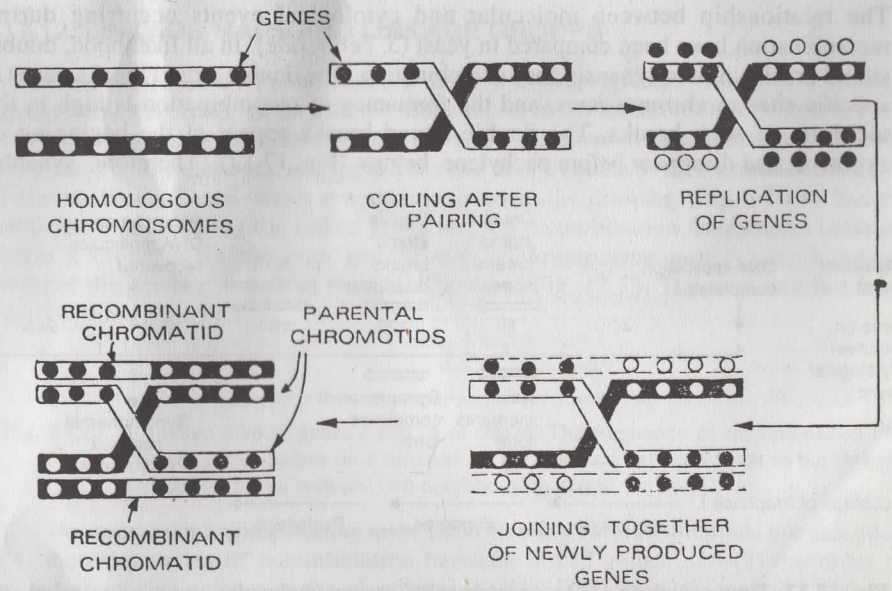


Fig. 17.10. The mechanism of crossing over proposed by Belling; this is the basis for the copy-choice theory of recombination.

newly synthesized. If one molecule serves as a template upto a certain distance and after that the homologous DNA molecule is used as the template, the newly produced DNA molecule will be recombinant (Fig. 17.11). This type of DNA replication is called ***copy-choice DNA replication.*** This theory also requires that (1) chromosome replication takes place after synapsis and that (2) DNA replication must be conservative. *These are contrary to the known facts: chromosome replication occurs before synapsis and DNA replication is universally semiconservative.* ***Therefore, copy-choice theory is virtually untenable.***

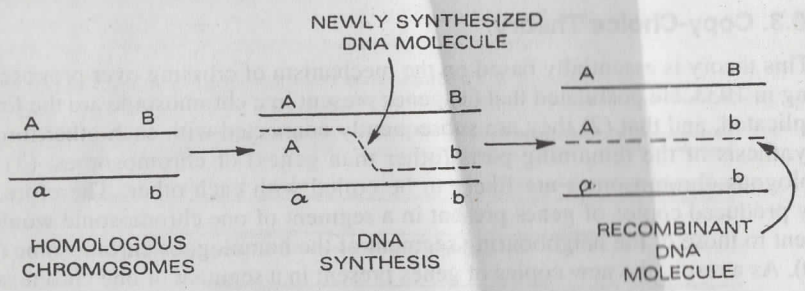


Fig. 17.11. Production of a newly synthesized recombinant DNA Molecule due to copy-choice DNA synthesis. In such a DNA synthesis, DNA molecule *A B* serves as the template upto a part of its length after which the DNA molecule *a b* is used as template. As a result, the newly synthesized molecule *A b* is a recombinant. For simplicity, each DNA molecule is represented by a single line.

17.11. MOLECULAR RECOMBINATION IN RELATION TO CYTOLOGICAL EVENTS

The relationship between molecular and cytological events occurring during recombination have been compared in yeast (*S. cerevisiae*). In all likelihood, double strand breaks initiate synapsis and recombination. The double strand breaks occur at specific sites in chromosomes, and the frequency of recombination is high in the vicitivity of such breaks. The double strand breaks appear at the beginning of zygotene, and disappear before pachytene begins (Fig. 17·12). Therefore, synapto-

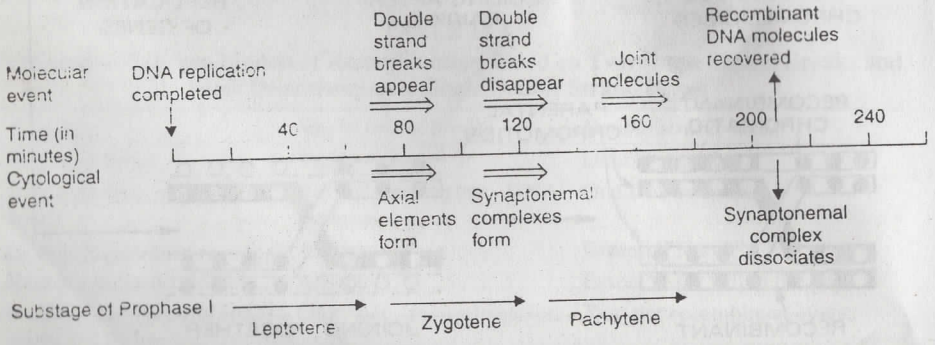


Fig. 17.12. Time sequence of the occurrence of various molecular and cytological events associated with recombination in yeast (*S. cerevisiae*). The time is measured from the end of premeiotic S phase.

nemal complex seems to the consequence and not the cause of crossing over. However, the formation of synaptonemal complex coincides with the presumed time of crossing over.

In yeast, recombinant DNA molecules appear at the end of pachytene, and recombination seems to be completed well after the formation of synaptonemal complex. At the beginning of diplotene, synaptonemal complex disappears and chiasmata become visible. The molecular nature of chiasmata remains unknown. *It is likely that chiasmata are some remnants of recombination, and not the actual intermediate stage of recombination. The occurrence of synaptonemal complex is correlated with that of recombination, but its role in recombination is not known.*

The two homologous chromosomes of a bivalent lie 200 nm away from each other during zygotenes and pachytene, and are separated from each other by the proteinacious synaptonemal complex. It is not known as to how recombination takes place in such a situation.

Linkage Map

- Depiction of the genes linked together on a straight line.
- Order of genes the same as in chromosome.
- Distance between genes in map units.
- Synonyms : genetic map, chromosomes map.
- Prepared using data from crosses involving 3 genes.
- *Map unit* : one map unit equals 1% recombination between linked genes; Syn., *cM* (centiMorgan).
- 1 map unit = ~1,000 kb DNA in man.

17.12. LINKAGE MAPS AND LINKAGE GROUPS

All the genes that are linked together form a ***linkage group.*** Genes of a linkage group can be represented on a single straight line in the same order in which they are normally present in the chromosome. In such a representation, the distance between two neighbouring genes is proportional to the frequency of recombination (%) between them. These values are also depicted in the drawing (Fig. 17.13). Such a line drawing depicting the linked genes and the recombination frequencies between them is known as ***linkage map, genetic map*** or ***chromosome map***. A simple linkage map of the genes *c* and *sh* in maize is shown in Fig. 17.13. The map unit has been derived form Fig. 17.1.

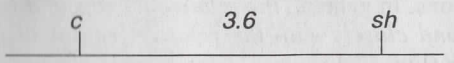


Fig. 17.13. A linkage map of genes *c* and *sh* of maize. The frequency of recombination (%) between genes serves as a measure of the distance between them in the linkage map and is shown between two neighbouring genes.

For preparing a chromosome map, the following two informations are essential: (1) the frequencies of recombination between linked genes, and (2) the order or sequence of these genes in the chromosome. ***Recombination frequencies between linked genes*** are determined from appropriate test crosses. These per cent frequencies are used as map units for preparing linkage maps. A ***map unit*** is that

distance in a chromosome, which permits one per cent recombination between two linked genes. A map unit is also called a ***centi-Morgan*** (*cM*). It may be noted that *map unit is an imaginary distance and it does not represent the actual physical distance between two linked genes in the chromosome*. The actual length of chromosome represented by a map unit varies considerably mainly depending on the organism in question (Table 17.2).

TABLE 17.2
The Physical Distance in Chromosomes Represented by 1 Map Unit

Organism	*Physical distance represented by 1 map unit*
Man	1,000 kb*
Mouse	1,500 kb

* kb, kilobase pairs.

The sequence of linked genes is determined by studying test crosses for three linked genes at a time. The data from such a test cross provide information on the order of the three genes in the chromosome as well as the frequency of recombination among them. To begin with, three linked genes are mapped. Subsequently, a three-point test cross involving any two of the three already mapped linked genes and a new gene expected to be linked with them is studied to map the new gene. In this manner, each new gene is mapped. It is desirable to include only those genes that show less than 20% recombination with each other.

Linkage Group
- All the genes located in the same chromosome.
- Their number always equals the gametic chromosome number (*n*).
- Their relative lengths correspond closely with those of the chromosomes in which they are located.
- *Genes located about 80 map units apart show independent segregation.*

The number of different linkage groups in a species is, as a rule, equal to its gametic chromosome number (n). For example, the number of linkage groups in *D.* melanogaster is four, it is 7 in barley, and 10 in maize. Each linkage group of a species is assigned to a specific chromosome of that species with the help of chromosomal aberrations. In general, *the relative lengths of different linkage groups of a species correspond closely with the relative lengths of the chromosomes in which they are located* (Fig. 17.11).

Linkage maps provide information on the genes that are linked together and the frequencies of recombination that may be expected between them. The total map distance between two genes present in a linkage group may exceed 50 or even 100 (Fig. 17.14). But it does not mean that such two genes would show more than 50% recombination. ***The frequency of recombination between two linked genes cannot exceed 50%, which is the frequency in case of independent segregation. Usually two linked genes would show independent segregation if they are more than about 80 map units apart.***

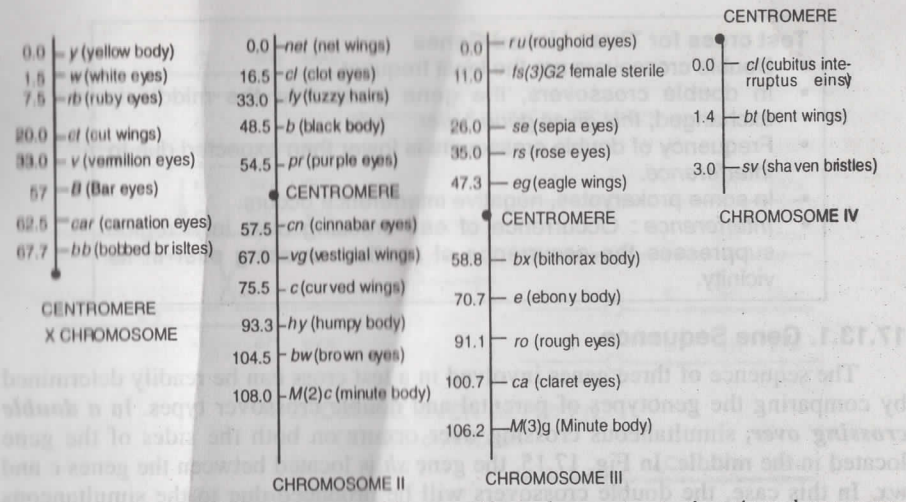


Fig. 17.14. The four linkage groups of *Drosophila melanogaster* (only selected genes are shown)

17.13. CROSSING OVER AMONG THREE LINKED GENES

A test cross for three linked genes, *e.g., c, sh,* and *wx* in maize, yields eight types of gametes and phenotypes (Table 17.3.). Two of these types (*e.g., C Wx Sh* and *c wx sh*) are the most frequent and represent the parental or nonrecombinant types. Two others (*e.g., CWx sh* and *cwx Sh*) are the least frequent and are double crossovers. The remaining four types (*e.g., C wx Sh, c Wx sh, c Wx Sh and C wx sh*) are produced by single crossing over between the three linked genes. ***As a rule, the parental types are the most frequent, the double cross-overs are the least frequent (or, sometimes, even absent), while the single crossovers are intermediate between these two.***

TABLE 17.3
The Different Types of Progeny Obtained from a Test-cross for Three Genes (*c, sh, wx*) in Maize

Genotype	*Phenotype**	*Number*	*Per cent frequency***
C Wx Sh/ c wx sh	Coloured, nonwaxy, full	2777	39.7
c wx sh/ c wx sh	Colourless, waxy, shrunken	2708	38.7
C wx sh/ c wx sh	Coloured, waxy, shrunken	116	1.7
c Wx Sh/ c wx sh	Colourless, nonwaxy, full	123	1.8
C wx Sh/ c wx sh	Coloured, waxy, full	643	9.2
c Wx sh/ c wx sh	Colourless, nonwaxy, shrunken	626	8.9
C Wx sh/ c wx sh	Coloured, nonwaxy, shrunken	4	0.06
c wx Sh/ c wx sh	Colourless, waxy, full	3	0.04
Total		7000	100

* Since the alleles *c wx sh*, uniformly contributed by the triple recessive test cross parent, have no effect on the phenotype, and are common to all the genotypes, (1) they may be ignored and (2) the phenotypes may be represented as '*C Wx Sh*' etc. in the place of 'coloured, nonwaxy, shrunken' etc., for brevity and simplicity.

** Per cent frequency = $\frac{\text{Number of progeny in a phenotypic class}}{\text{Total number of progeny}} \times 100$

Test cross for Three Linked Genes

- Double crossovers are the least frequent.
- In double crossovers, the gene located in the middle is exchanged; *this gives gene order.*
- Frequency of double crossovers is lower than expected due to *interference.*
- In some prokaryotes, negative interference occurs.
- *Interference* : Occurrence of one crossing over in a region suppresses the occurrence of another crossing over in its vicinity.

17.13.1. Gene Sequence

The sequence of three genes involved in a test cross can be readily determined by comparing the genotypes of parental and double crossover types. In ***a double crossing over,*** simultaneous crossing over occurs on both the sides of the gene located in the middle. In Fig. 17.15, the gene *sh* is located between the genes *c* and *wx*. In this case, the double crossovers will be produced due to the simultaneous occurrence of crossing over between genes *c* and *sh* and between *sh* and *wx*.

In this figure, the parental type chromosomes (or gametes or phenotypes) are *C Sh Wx* and *c sh wx*, while the double crossover chromosomes are *C sh Wx* and *c Sh wx*. A comparison of the parental and the double crossover types shows that *in the double crossovers only the gene located in the middle* (sh) *has switched its position.* This property of the double crossovers is used for determining the gene located in the middle.

In table 17.3, the parental types are *C Wx Sh* and *c wx sh*, whereas the double crossovers are *C Wx sh* and *c wx Sh*. Here the alleles of gene *sh* have switched their position in the double crossovers as compared to the parental types. Therefore, the gene *sh* must be located between the genes *c* and *wx*. This is the sequence of these three genes in the chromosome.

17.13.2. Recombination Frequencies

In Table 17.3, the parental types are *C Wx Sh* and *c wx sh*. The gene *sh* is located in the middle. Therefore, we estimate the recombination frequencies between the genes *c* and *sh*, and between *sh* and *wx*. These two estimates will allow us to prepare the genetic map using the gene sequence deduced above.

Recombination between genes *c* and *sh* would yield the types *C wx sh*, *c Wx Sh* and the double crossovers. In these classes, *C* is present with *sh* and *c* is present with *Sh* as compared to the parental combinations *C-Sh* and *c-sh*. Hence,

Recombination (%) between *c* and *sh* = Sum of the frequencies (%) of *C wx sh*, *c Wx Sh*, and the double crossovers
= 1.7 + 1.8 + 0.06 + 0.04 per cent
= 3.6 per cent

Similarly, recombination between the genes *wx* and *sh* will give the types *C wx Sh*, *c Wx sh* and the double crossovers. In these classes, the combinations *wx Sh* and *Wx sh* are present as against the parental combinations *Wx Sh* and *wx sh*. Therefore,

Recombination (%) between *sh* and *wx* = Sum of the frequencies (%) of *C wx Sh*, *cWx sh* and the double crossovers
= 9.2 + 8.9 + 0.06 + 0.04 per cent
= 18.2 per cent.

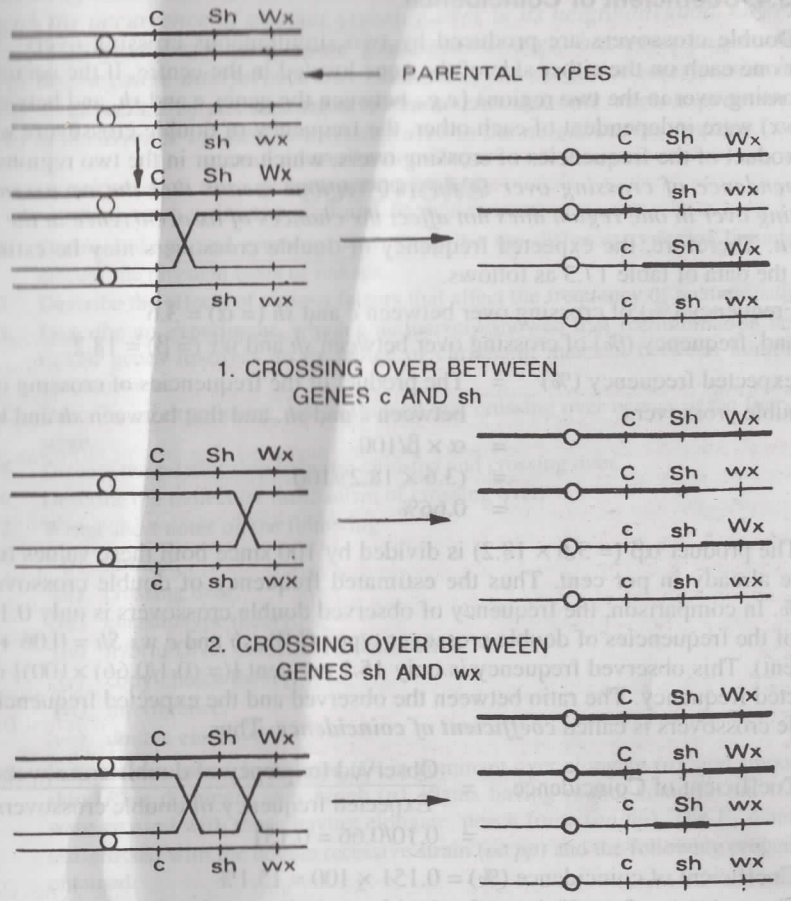


Fig. 17.15. Crossing over in the different regions between three linked genes *c*, *sh* and *wx*, and their consequences: (1) crossing over between genes *c* and *sh*, (2) crossing over between *sh* and *wx*, and (3) simultaneous crossing over between genes *c* and *sh*, and between *sh* and *wx* (double crossing over). Note that in double crossing over the position of only the gene located in the middle changes.

17.13.3. Linkage Map

We can now map the genes *c*, *sh* and *wx* in the order *c-sh-wx* at the distance of 3.6 and 18.2 map units, respectively (Fig. 17.16). A new gene expected to be linked with these genes can be mapped by ordering the new gene with either genes *c* and *sh* or *sh* and *wx* in a three point test cross.

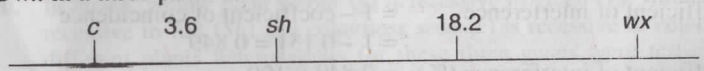


Fig. 17.16. A linkage map of the genes *c*, *sh* and *wx* of maize.

17.13.4. Coefficient of Coincidence

Double crossovers are produced by two simultaneous crossing overs. These occur one each on the either side of the gene located in the centre. If the occurrence of crossing over in the two regions (*e.g.*, between the genes *c* and *sh*, and between *sh* and *wx*) were independent of each other, the frequency of double crossovers will be the product of the frequencies of crossing overs, which occur in the two regions. *The independence of crossing over in the two regions means that the occurrence of crossing over in one region does not affect the chances of its occurrence in the other region*. Therefore, the expected frequency of double crossovers may be estimated from the data of table 17.3 as follows.

Frequency (%) of crossing over between *c* and *sh* (= α) = 3.6

and, frequency (%) of crossing over between *sh* and *wx* (= β) = 18.2

The expected frequency (%) of double crossovers = The product of the frequencies of crossing overs between *c* and *sh*, and that between *sh* and *wx*

$= \alpha \times \beta/100$

$= (3.6 \times 18.2)/100$

$= 0.66\%$

The product αβ (= 3.6 × 18.2) is divided by 100 since both these values (α and β) are already in per cent. Thus the estimated frequency of double crossovers is 0.66%. In comparison, the frequency of observed double crossovers is only 0.1% (= sum of the frequencies of double crossover types *C Wx sh* and *c wx Sh* = 0.06 + 0.04 per cent). This observed frequency is only 15.1 per cent [(= (0.1/0.66) ×100)] of the expected frequency. The ratio between the observed and the expected frequencies of double crossovers is called ***coefficient of coincidence***. Thus,

$$\text{Coefficient of Coincidence} = \frac{\text{Observed frequency of double crossovers}}{\text{Expected frequency of double crossovers}}$$

$= 0.10/0.66 = 0.151$

and, Coefficient of coincidence (%) = 0.151 × 100 = 15.1%

The estimate of coefficient of coincidence indicates the degree of agreement between the observed and the expected frequencies of double crossovers.

17.13.5. Coefficient of Interference

As a rule, the observed frequencies of double crossovers are lower than their expected values. This is interpreted as follows: *the occurrence of crossing over in one region of a chromosome interferes with its occurrence in the neighboring segments; this is called* ***interference.*** It may be expected that the intensity of interference would progressively decrease as the point of second crossing over becomes farther from that of the first one. Therefore, the coefficient of coincidence would be lower when the concerned genes are located close to each other than when they are located farther apart. The intensity of interference may be estimated as ***coefficient of interference,*** which is equal to one minus the coefficient of coincidence. Thus,

Coefficient of interference = 1 – coefficient of coincidence

$= 1 - 0.151 = 0.849$

and, Coefficient of interference (%) $= 0.849 \times 100$

$= 84.9\%$

The coefficient of interference indicates the degree of interference by a crossing over with the occurrence of another crossing over in its neighbourhood. Generally, it is a positive value, but in some prokaryotes, e.g., *bacteriophages and in* Aspergillus, *negative interference has been observed. Negative interference means that the occurrence of a crossing over in a chromosome region promotes the chances of occurrence of another crossing over in the neighbouring regions.*

QUESTIONS

1. Define linkage and crossing over. Describe the different types of linkage, and discuss the physical basis of linkage.
2. Describe the effects of various factors that affect the frequency of recombination.
3. Describe an experiment, which conclusively showed that recombination between linked genes involves an exchange of chromatin material between homologous chromosomes.
4. Describe an experiment, which showed that crossing over occurs in the four-strand stage.
5. Discuss the relationship between chiasma and crossing over.
6. Describe the molecular mechanism of crossing over.
7. Writer short notes on the following:
   (*i*) linkage,
   (*ii*) crossing over,
   (*iii*) chiasmatype theory,
   (*iv*) hybrid DNA model of recombination,
   (*v*) copy-choice theory,
   (*vi*) interference,
   (*vii*) double crossovers.
8. In tomatoes, round fruit shape (*O*) is dominant over elongate (*o*), and smooth fruit skin (*P*) is dominant over peach (*p*). Plants having round, smooth fruit (*OO PP*) were crossed with those having elongate, peach fruits (*oo pp*). The F_1 plants were test-crossed with the double recessive strain (*oo pp*) and the following progeny were obtained.

Round, smooth	*Round, peach*	*Elongate, smooth*	*Elongate, peach*
(Oo Pp)	*(Oo pp)*	*(oo Pp)*	*(oo pp)*
360	*32*	*38*	*370*

   (*i*) Test if the genes *o* and *p* are linked. If so, in which phase ?
   (*ii*) Estimate the frequency of recombination between the genes *o* and *p* and prepare their genetic map.
9. Tomato plants having round, peach fruits were crossed with those having elongate, smooth fruits; the F_1 plants were test-crossed to yield the following data.

Round, smooth	Round, peach	Elongate, smooth	Elongate, peach
14	133	141	12

   (*i*) Test if the genes for round and peach fruits are linked. If so, in which phase ?
   (*ii*) Estimate the frequency of recombination between the two genes and prepare a linkage map for them.
10. In maize, yellow plant colour (*y*) is recessive to green (*Y*), shrunken seed (*sh*) is recessive to full (*Sh*) and colourless seed (*c*) is recessive to coloured (*C*). Three different plants heterozygous for these three genes were test-crossed and the following data were obtained.

Phenotype	Progeny from the heterozygous plant		
	I	*II*	*III*
Y C Sh	368	22	95
y c sh	387	23	100
y C Sh	21	390	3
Y c sh	24	365	2
y c Sh	98	0	375
Y C sh	97	5	380
y C sh	4	96	20
Y c Sh	1	99	25
Total	1,000	1,000	1,000

(*i*) Identify the parental and double crossover types in the above three cases.
(*ii*) Determine the sequence of the three genes.
(*iii*) Estimate the frequencies of recombination between these genes and prepare their linkage map.

11. In tomato, compound inflorescence (*s*) is recessive to solitary (*S*). When plants heterozygous for the genes *o, p* and *s* were test-crossed, the following data were obtained.

OPS	73
ops	63
OPs	348
opS	306
Ops	96
oPS	110
OpS	2
oPs	2

From the above data, determine the following:

(*i*) genotypes of the two parents,
(*ii*) genotype and phenotype of F_1,
(*iii*) different gametes produced by F_1,
(*iv*) frequency of recombination between the genes *o, p* and *s*, and
(*v*) prepare their genetic map.

12. In *Drosophila, cinnabar* eye colour (*cn*) is recessive to the normal red colour, and *vestigial* wings (*vg*) are recessive to normal wings. When females heterozygous for *cinnabar* and *vestigial* wing genes are mated with males having *cinnabar* eyes and *vestigial* wings, the following progeny were obtained.

Normal (= red eye, normal wings)	Cinnabar (= cinnabar eye, normal wings)	Vestigial (= red eye, vestigial wings)	Cinnabar, vestigial
493	53	48	476

(*i*) Are the two genes linked ?
(*ii*) If so, in which phase
(*iii*) Estimate the frequency of crossing over between them and
(*iv*) prepare a linkage map of the two genes.

13. When some other females heterozygous for the *cinnabar* and *vestigial* genes were mated with *cinnabar, vestigial* males, them following progeny were obtained.

Normal	Cinnabar	Vestigial	Cinnabar, vestigial
39	328	313	29

(*i*) Explain these results and compare them with those in question 12. (*ii*) Are these genes linked and, if so, in which phase ? (*iii*) Determine the frequency of recombination between the two genes. (*iv*) Prepare a linkage map for these genes.

14. In *Drosophila*, genes *rs* (*rose* eye colour) and *eg* (*eagle* wings) are located about 12 crossover units apart in chromosome III. Males heterozygous for these genes are mated with *rose* eye, *eagle* wing females. Give the genotypes, phenotypes and per cent frequencies of the different types of progeny expected from these matings. Give the reasons in support of your answer.

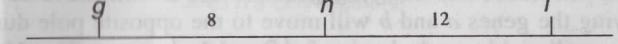
15. Give the genotypes, phenotypes and per cent frequencies of the different types of progeny obtained from matings between *Rs rs Eg eg Drosophila* females and *rs rs eg eg* males. (Genes *rs* and *eg* are about 12 crossover units apart in chromosome III.)

16. In *Drosophila, black* body (*b*) is recessive to normal gray body, *purple* eyes (*pr*) is recessive to red eyes and *vestigial* wings is recessive to normal wings. When females heterozygous for these three genes were mated with the triple recessive males (*bb vv prpr*) the following progeny were obtained.

Normal	824
black, vestigial, purple	796
black	65
vestigial, purple	55
vestigial	112
black, purple	138
purple	7
black, vestigial	3
Total	2000

(*i*) Identify the parental and double crossover types. (*ii*) Determine the gene sequence, (*iii*) the frequencies of crossing over between the genes, (*iv*) the coefficients of coincidence and interference, and (*v*) prepare their linkage map.

17. The linkage map of three genes is given below. The heterozygote *GHI/ghi* is crossed with *ghi/ghi*. If 2,000 progeny were obtained from this cross, determine the frequencies of the different types of progeny. The coefficient of interference is 40%.



18. In *Drosophila, cubitus interrupts* veins (*ci*) are recessive to normal veins, *bent* wings (*bt*) are recessive to normal wings and *shaven* bristles (*sh*) are recessive to normal bristles. Females heterozygous for these genes were mated with *cubitus interrupts, bent* wings and *shaven* bristles males. The following progeny were obtained.

Normal	956
Cubitus interruptus, bent, shaven	984
Cubitus interruptus	12
Bent, shaven	16
Shaven	17
Cubitus interruptus, bent	15

(**Hint :** This is a three point test cross. Therefore, eight phenotypic classes are expected. Obviously, two phenotypic classes are absent.)

19. *Drosophila* males heterozygous for the genes *cubitus interruptus, bent* wings and *shaven* bristles are mated with females having *cubitus interruptus* veins, *bent* wings and *shaven* bristles. If this mating produced 10,000 flies, give the expected frequencies of the different phenotypic classes.

18

Recombination in Fungi and Bacteria

18.1. INTRODUCTION

Generally, fungi are haploid (*n*); hence the products of meiosis of zygotes can be directly analysed and there is no need for the test-crosses so common in the case of diploid organisms. In many fungi, the spores produced due to different meiotic events are enclosed in separate membranes, *e.g.*, in asci in *Neurospora*. These structures are called ***tetrads*** since each of them has the four haploid cells, called a tetrad, resulting from one complete meiotic division. In some fungi, *e.g.*, *Neurospora*, the products of different divisions during spore formation are arranged in the precise sequence in which they are produced (***ordered tetrads***); this permits the estimation of recombination frequency between a gene and its centromere. But in many other fungi, *e.g.*, yeast, the spores are not present in the same sequence in which they were produced (***unordered tetrads***); in such cases, the frequency of recombination between a gene and its centromere cannot be determined.

18.2. TETRAD ANALYSIS

In fungi, the detection of linkage and the estimation of recombination frequencies between genes is based on the classification of a large number of tetrads produced by the appropriate heterozygotes into different classes; this is known as ***tetrad analysis***. Suppose, strains having genotypes *AB* and *ab*, are crossed to obtain *Aa Bb* zygotes, and the two genes, *Aa* and *Bb*, are segregating independently. In many zygotes, the chromosomes carrying genes *A* and *B* will move to one pole, while those carrying the genes *a* and *b* will move to the opposite pole during anaphase I; such zygotes will yield tetrads having 2 *AB* and 2 *ab* spores (Fig. 13·5). Such tetrads are known as ***parental ditype (PD)***, since they have two types of spores, which have the same genotypes as the two parents involved in the cross. In many other zygotes, the chromosomes having genes *A* and *b* will move to one pole, and those having *a* and *B* to the opposite pole during anaphase I. Tetrads from such zygotes will have two spores with the genotypes *Ab* and two with *aB* (Fig. 13·5). These tetrads are called ***nonparental ditype (NPD)*** as they have two types of spores having genotypes different from the two parents. These two types of tetrads (PD and NPD) will be equally frequent and will constitute the bulk of tetrads since they would result from an independent assortment of the two concerned chromosome pairs.

In practice, some tetrads contain four different types of spores representing all the four of the above genotypes (*AB*, *ab*, *aB* and *Ab*). Such tetrads are called ***tetratypes (T)***, since they contain four different types of spores. Tetratypes, in this case, are produced when a crossing over occurs either between the gene *A* and its centromere or between the gene *B* and its centromere (Fig. 18.1). It should be noted that when there is no crossing over between a gene and its centromere, segregation

for that gene occurs during the first anaphase (***first division segregation***, Fig. 13.6). But whenever crossing over occurs between a gene and its centromere, it segregates during the second anaphase (***second division segregation***, Fig. 13·7 and 18·1). Since tetratype tetrads are produced due to crossing over, their frequency would be relatively smaller than those of parental ditype and nonparental ditype tetrads.

Thus, *if two genes are segregating independently, parental ditype and nonparental ditype tetrads will be the most frequent and their frequencies will be comparable, while tetratype tetrads will be relatively infrequent.*

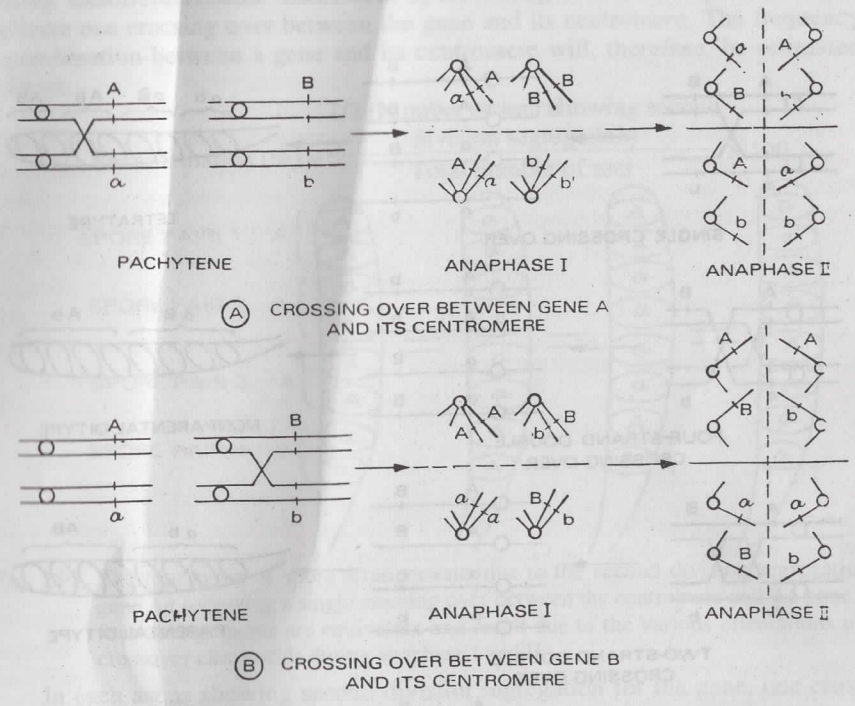


Fig. 18·1. Production of tetratype tetrads in the case of two genes (*Aa* and *Bb*) segregating independently. Such tetrads may originate either due to a crossing over between the gene *Aa* and its centromere or between the gene *Bb* and its centromere.

18.3. DETECTION OF LINKAGE

In fungi producing tetrads, linkage between genes is detected by comparing the relative frequencies of parental ditype, nonparental ditype and tetratype tetrads with those expected for independent segregation. *If the two genes are linked, parental ditype tetrads will be the most frequent, nonparental ditype ones will be the least frequent, while tetratype tetrads will have an intermediate frequency. In contrast, independent segregation produces equal frequencies of parental and nonparental ditype tetrads and a relatively low frequency of tetratype tetrads.*

The deviation due to linkage occurs because the production of parental ditype tetrads involves no crossing over (Fig. 18.1), tetratype tetrads are produced due to a single crossing over involving two chromatids (of the concerned bivalent; between the two linked genes), while nonparental ditype tetrads are the consequence of four-

strand double crossing over (one crossing over in each of the two pairs of nonsister chromatids of the concerned bivalent between the two linked genes; Fig. 18.2). As a result, the parental ditype tetrads are the most frequent (noncrossover), tetratype tetrads are less frequent than the former (one crossing over needed), while the nonparental ditypes are the least frequent (two simultaneous crossing overs involving all the four chromatids of the bivalent are needed). It may be noted that parental ditype tetrads will also be produced due to (two chromatid) double crossing over between the two genes. Similarly, a three-strand (involving three chromatids) double-crossing over also will generate tetratype tetrads (Fig. 18·2).

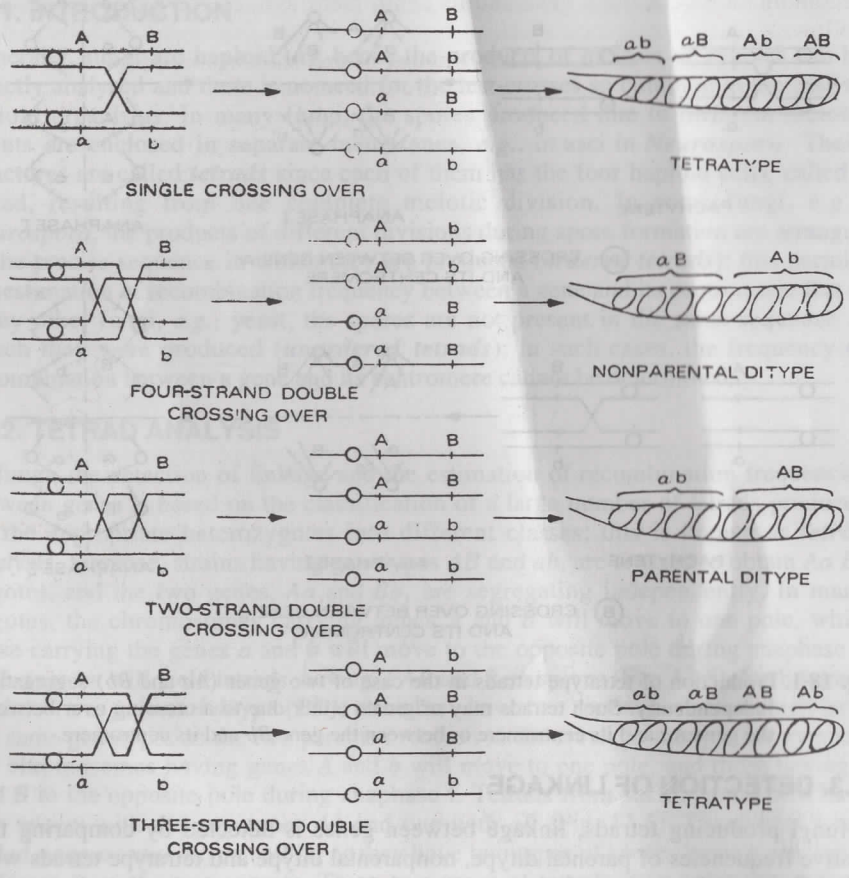


Fig. 18·2. Consequences of single crossing over, four-strand double crossing over, two-strand double crossing over and three-strand double crossing over between two genes *A* and *B* on the type of tetrad (ascus) produced.

18.4. ANALYSIS OF ORDERED TETRADS

Analysis of ordered tetrads permits the determination of recombination frequency between a gene and its centromere as well as that between linked genes.

18.4.1. Recombination between A Gene and Its Centromere

In *Neurospora*, a 4 : 4 spore arrangement is obtained if a crossing over does not occur between a gene, say, *Aa* and its centromere. In such a case, this gene would show first division segregation. But when a crossing over occurs between the gene and its centromere, it shows second division segregation and a $2A : 2a : 2A : 2a$ spore arrangement (Fig. 18.5). Second division segregation of the gene may also give rise to $2a : 2A : 2a : 2A$, $2A : 4a : 2A$ or $2a : 4A : 2a$ spore arrangements (Fig. 18.3) depending upon the orientation of the crossover and noncrossover chromatids during meiotic divisions. These four spore arrangements are equivalent as they indicate one crossing over between the gene and its centromere. The frequency of recombination between a gene and its centromere will, therefore, be estimated as follows.

$$\text{Recombination (\%)} = \frac{\text{1/2 (Number of asci showing second division segregation)}}{\text{Total number of asci}} \times 100$$

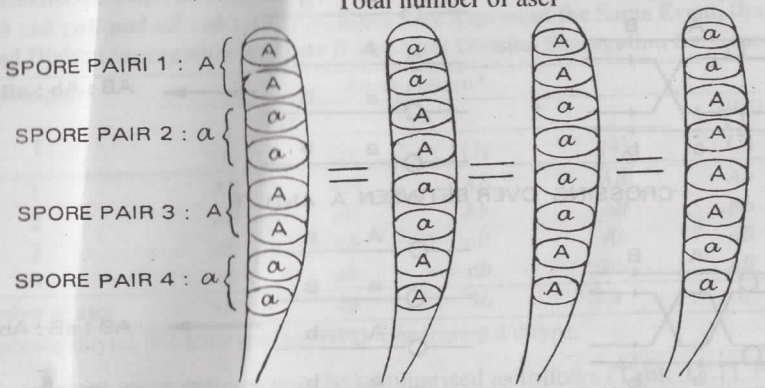


Fig. 18·3. The four types of spore arrangements due to the second division segregation of gene *Aa* following a single crossing over between the centromere and the gene. The four arrangements are equivalent and result due to the various orientations of the crossover chromatids during anaphase I and II.

In each ascus showing second division segregation for the gene, one crossing over would have occurred between the gene and its centromere. Consequently, only two of the four chromatids of the concerned bivalent will be crossover, the remaining two being noncrossover. That is, only one-half of the spores of such asci will be recombinant. For this reason, the number of asci showing second division segregation is multiplied by one-half while estimating the recombination frequency. For the same reason, the number of non-parental ditype tetrads is not multiplied by one-half, since all their products are recombinants (Fig. 18.1).

18.4.2. Recombination between Two Genes and Their Centromeres

Analysis of ordered tetrads permits the mapping of two linked genes relative to their centromeres. Crossing overs between the two genes and between each gene and its centromere produces a variety of spore arrangements in ordered tetrads depending upon (1) the gene arrangement, (2) the type of crossing over event and (3) the orientation of crossover and noncrossover chromatids during meiosis.

The two linked genes, say *A* and *B*, and the centromere may be present in the following three orders: Centromere-*A*-*B*, centromere-*B*-*A* and *A*-centromere-*B*. If the sequence were centromere-*A*-*B* and the linkage was in coupling phase (*AB/ab*), the

following spore arrangements would be produced due to crossing overs in the different regions (Fig. 18.4; Table 18.1). As single crossing over between the centromere and *A* will yield the ascus patterns: *AB : ab : AB : ab, ab : AB : ab: AB, AB : ab : ab : AB* and *ab : AB : AB : ab*. All these patterns show second division segregation for both the genes *A* and *B*, are parental ditype, and can be summarised as *AB : ab : AB : ab* (Table 18.1).

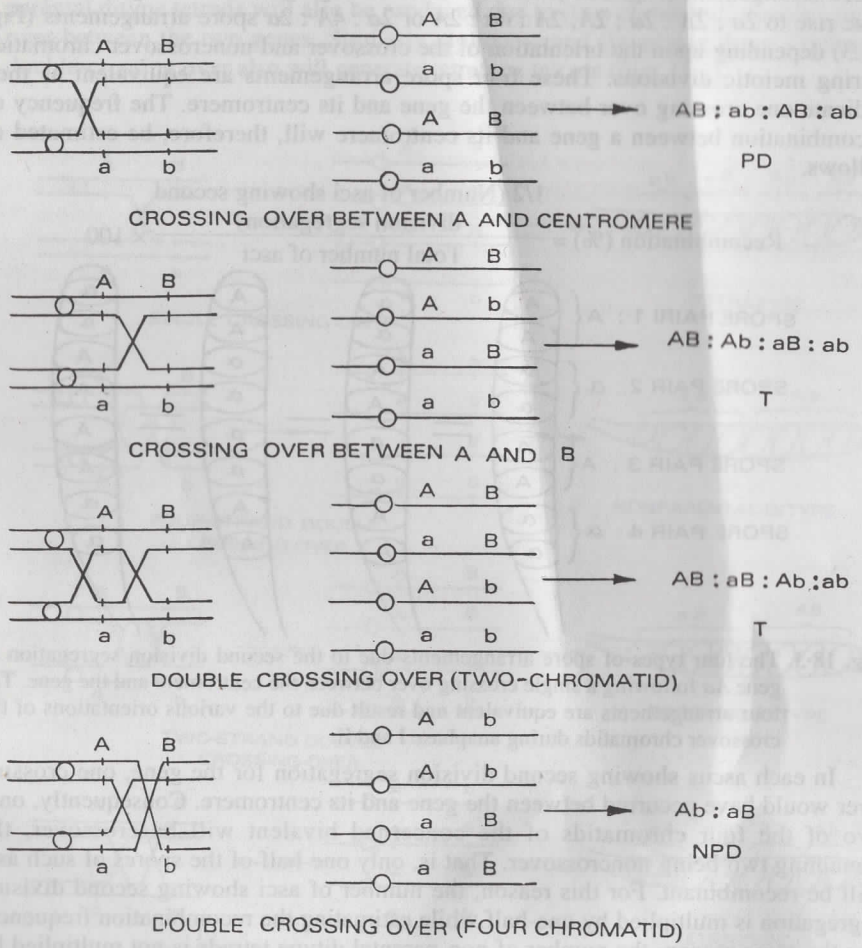


Fig. 18·4. Some of the main crossover events between linked genes *A* and *B* and their centromeres, and the spore arrangements generated by them (PD = parental ditype, T = tetratype, and NPD = nonparental ditype tetrads). Other crossing over events, *e.g.*, three-chromatid double and four-chromatid double crossing overs between centromere and *B*, and between *A* and *B* are possible but are not shown here; the consequences of these events can be easily worked out.

Similarly, a single crossing over between *A* and *B* will generate the following ascus patterns: *AB : Ab : AB : ab* (Fig. 18.4), *ab : aB : Ab : AB, AB : Ab : ab : aB, aB : ab : Ab : AB, Ab : AB : aB : ab, ab : aB : AB : Ab, Ab : AB : ab : aB* , and *aB : ab : AB : Ab*. All these pattern show second division segregation for *B* and first

division segregation for *A*, are tetratype, and can be summarised as *AB : Ab : aB : ab* (Table 18.1).

A two chromatid double crossing over (between centromere and *A*, and between *A* and *B*) would yield the spore patterns *AB : aB : Ab : ab* (Fig. 19.4), *ab : Ab : aB : AB, aB : AB : Ab : ab, ab : Ab : AB : ab, AB : aB : ab : Ab, Ab : ab : aB : AB, aB : AB : ab : Ab* and *Ab : ab : AB : aB*. All these patterns show second division segregation for *A*, are tetratype, and can be summarised as *AB : aB : Ab : ab* (Table 18.1). But a four-chromatid double crossing over between genes *A* and *B* will yield the spore arrangement *Ab : Ab : aB : aB*, which is nonparental ditype.

TABLE 18.1

Frequencies of Different Spore Arrangements in Asci Produced by the Heterozygotes *AB ab* (Derived from the Cross *AB* × *ab*) of *Neurospora*. Each Asucs Arrangement Includes all Ascus Patterns Possible for the Arrangement, *i.e.*, Pattern (3) Includes the Arrangements *AB : Ab : aB : ab, ab : aB : Ab : AB, Ab : AB : aB : ab, ab : aB : AB : Ab, AB : Ab : ab : aB* and *aB : ab : Ab : AB*, Since they Represent the Same Event, that is, Second Divison Segregation for Gene *B* and First Division Segregation for Gene *A*

Spore pair	*Ascus pattern**				
	PD		*T*		*NPD*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
1	*AB*	*AB*	*AB*	*AB*	*Ab*
2	*AB*	*ab*	*Ab*	*aB*	*Ab*
3	*ab*	*AB*	*aB*	*Ab*	*aB*
4	*ab*	*ab*	*ab*	*ab*	*aB*
Number of asci	312	48	36	4	0

*PD = parental ditype, T = tetratype, and NPD = nonparental ditype.

The different ascus patterns may be summarised as follows (Table 18.1). Pattern (1), *i.e., AB : AB : ab : ab*, is parental ditype, involves no crossing over and will be the most frequent. Patterns (4), *viz, AB : aB : Ab : ab*, and (5), *i.e., Ab : Ab : aB : aB*, will be produced due to double crossing overs, and will be the least frequent. But asci *AB : ab : AB : ab* [pattern (2)] and *AB : ab : aB : ab* [pattern (3)] result from single crossing overs between centromere and *A* and between *A* and *B*, respectively, and will have intermediate frequencies.

It can be seen that if the gene order were centromere-*B*-*A*, pattern (3), *viz., AB : Ab : aB : ab*, can result only due to two-chromatid double crossing over, hence it would be the least frequent. In contrast, the pattern (4), *AB : aB : Ab : ab*, will be produced due to single crossing over between *B* and *A*; therefore, it would have intermediate frequency. However, if the gene order were *A*-centromere-*B*, ascus pattern (2), *AB : ab : AB : ab*, will not be produced, while patterns (3), *AB : Ab : aB : ab*, and (4), *AB : aB : Ab : ab*, will exhibit intermediate frequencies (depending on the distances of genes *A* and *B* from the centromere). Thus, a comparison of the frequencies of ascus patterns (3), *AB : Ab : aB : ab*, and (4), *AB : aB : Ab : ab*, will yield the order of the two linked genes. Another way for deducing the gene order is to apply the following criterion: *whenever the gene located closer to centromere shows second division segregation, the gene located farther from the centromere also shows second division segregation (except in the case of two-chromatid double crossing over). But the distal gene will show second division segregation without the proximal gene doing so whenever* **crossing over** *occurs between the two genes.*

In Table 18.1, ascus pattern (2), *AB : ab : AB : ab*, is present, pattern (3), *AB : Ab : aB : ab* , is intermediate and pattern (4), *AB : aB : Ab : ab*, is the least frequent. Therefore, the gene order would be centromere-*A*-*B*. Ascus patterns (2) and (4) are produced due to crossing over between centromere and *A*, and two-chromatid double crossing over, respectively. Therefore, the frequency of recombination between centromere and *A* will be.

$$= \frac{1/2 \text{ (Number of asci with (2) and (4) patterns)}}{\text{Total number of asci}} \times 100$$

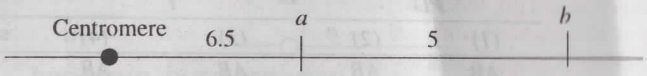
$$= \frac{1/2\ (48 + 4)}{400} \times 100 = 6.5\% \text{ (or map units)}$$

Similarly, ascus pattern (3) is produced due to a single crossing over between *A* and *B*. Therefore, the frequency of recombination between *A* and *B* will be,

$$= \frac{1/2 \text{ (Number of asci with (3) and (4) patterns)}}{\text{Total number of asci}} \times 100$$

$$= \frac{1/2\ (36 + 4)}{400} \times 100 = 5.0\% \text{ (or map units)}$$

The genes *A* and *B* and their centromere can now be mapped as follows.



18.5. ANALYSIS OF UNORDERED TETRADS

The determination of ascus patterns is tedius and time consuming involving the dissection of individual asci. It is much easier to allow individual asci to dehisce and analyse them as unordered tetrads. In addition, some Ascomycetes, *e.g.*, yeast, and algae, *e.g.*, *Chlamydomonas*, produce unordered tetrads. Analysis of such tetrads does not permit the estimation of recombination frequency between a gene and its centromere. But the frequency of recombination between two linked genes can be readily determined as follows.

$$\text{Recombination frequency (\%)} = \frac{1/2 \text{ (number of tetratype asci)} + \text{number of nonparental ditype asci}}{\text{Total number of asci}} \times 100$$

Unordered tetrad data have been of great value in the study of nonreciprocal recombination or gene conversion.

18.6. GENE CONVERSION

Tetrad analysis in Ascomycetes has enabled discovery of the phenomenon of ***nonreciprocal recombination*** for which the most commonly used term is ***gene conversion***. When a pair of alleles, say m^+/m, is segregating, tetrads are expected to shown $4m^+ : 4m$ or $2m^+ : 2m$ spores. But occasional tetrads (about 4 per thousand) show a deviation from this expectation; they may show $6m^+ : 2m$, $5m^+ : 3m$ or even $7m^- : 1m$ spore ratios. In such cases ($6m^+ : 2m$ etc.), it appears as if some of the *m* alleles got converted into the m^+ allele leading to an excess of m^+ at the cost of *m* spores. This was the reason for the name ***gene conversion*** for this phenomenon.

Gene conversion is not due to mutation of one allele of a gene to the other since it occurs at much higher rates than the corresponding mutations. It has been explained on the basis of nonreciprocal recombination as it always produces the

allele present on the homologous chromosome, and is correlated about 50% of the time with reciprocal recombination of markers located on either side of the 'converted' gene. Thus gene conversion is correlated with crossing over, but its occurrence does not depend on that of the latter.

18.7. RECOMBINATION DURING MITOSIS

Crossing over occurs during mitosis leading to recombination between linked genes in somatic tissues. Somatic or mitotic crossing over was first inferred in 1931 by Stern from the production of twin spots in *Drosophila* flies heterozygous for the sex-linked genes *yellow* body colour and *singed* (bent blunted) bristles present in repulsion phase (*y Sn/Y sn*). These flies will have grey body and normal bristles. A twin spot in them will consist of a spot of yellow colour and a neighbouring spot with singed bristles. The origin of such twin spots can be readily explained by a crossing over between the centromere and the gene nearest to it (Fig. 18.5).

Data on mitotic recombination are used for chromosome mapping in some fungi, *e.g.*, *Aspergillus nidulans*. In *Aspergillus* and many other fungi, vegetative hyphae of different genotypes often fuse to produce dicaryotic cells called ***heterocaryons***; often, the two haploid nuclei fuse to yield diploid ($2n$) hyphae. The diploid hyphae may undergo ***haploidization*** (not involving meiosis), which may occassionally be accompanied with crossing over. But crossing over occurs more frequently in the diploid stage and data from them are used for the determination of recombination frequencies.

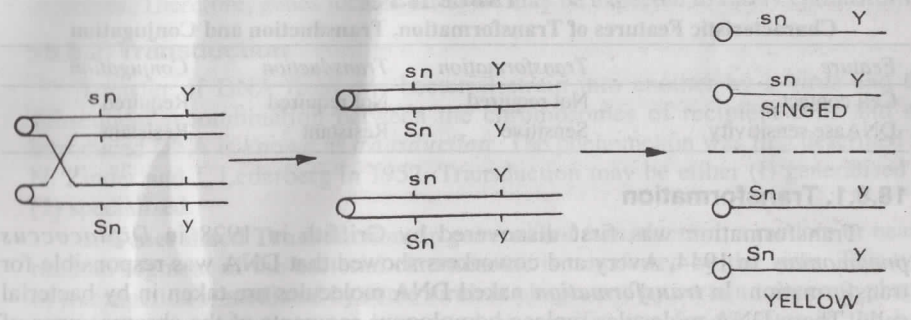


Fig. 18·5. Crossing over between *sn* and its centromere leading to the production of one daughter cell homozygous for *sn* (giving rise to *singed* spot) and another homozygous for *y* (generating the neighbouring *yellow* spot). Such cells give rise to twin spots.

Suppose a diploid strain is heterozygous for three genes *a, b, c* linked in the coupling phase ($abc/a^+b^+c^+$) and the gene sequence is centromere *–a-b-c*. A single crossing over between centromere and *a* will yield the twin homozygotes *aa bb cc* and $a\ ^+a^+\ b^+b^+\ c^+c^+$; that between *a* and *b* will give $a^+a\ bb\ cc$ and $a^+a\ b^+b^+c^+c^+$; while that between *b* and *c* will generate $a^+a\ b^+b\ cc$ and $a^+a\ b^+b\ c^+c^+$. The consequences of double and triple crossing overs can be readily worked out, but they are not considered here as they will be relatively infrequent. In this case (gene order, centromere—*a-b-c*), *a homozygote for gene* a *will almost always be homozygous for genes* b *and* c. *Similarly, individuals homozygous for* b *will almost always be homozygous for* c, *but they may or may not be homozygous for* a. *Finally, those homozygous for* c *may or may not be homozygous for* a *or* b. This expectation will

not be realized only in the case of double and triple crossovers, which are expected to be infrequent.

The expectation for other sequences of the centromere and genes *a*, *b* and *c* may be easily worked out. The actual data are compared with these expectations to determine the gene and centromere sequence and for the estimation of recombination frequencies between them.

18.8. RECOMBINATION IN BACTERIA

Recombination in bacteria may occur through the following modes: (1) transformation, (2) transduction and (3) conjugation. These three modes can be differentiated on the basis of necessity for cell contact and their sensitivity to DNAase (deoxyribonuclease). Both transformation and transduction do not depend on cell contact, but transformation is DNAase sensitive, while transduction is not. In contrast, conjugation requires cell contact and is DNAase resistant (Table 18.2). These two criteria may be easily tested by placing the cultures of two different genotypes of a bacterium in the opposite arms of a U-tube. The two cultures are separated by a filter, which permits the movement of DNA and virus particle but not of bacterial cells. After some time, the bacterial cultures are tested for the presence of recombinant cells. This experiment will test the necessity for cell contact. In another similar experiment, DNAase may be added to the cultures to test the sensitivity of recombination process to DNAase.

TABLE 18.2
Characteristic Features of Transformation, Transduction and Conjugation

Feature	*Transformation*	*Transduction*	*Conjugation*
Cell contact	Not required	Not required	Required
DNAase sensitivity	Sensitive	Resistant	Resistant

18.8.1. Transformation

Transformation was first discovered by Griffith in 1928 in *Diplococcus pneumoniae*. In 1944, Avery and coworkers showed that DNA was responsible for transformation. In ***transformation*** naked DNA molecules are taken in by bacterial cells. These DNA molecules replace homologous segments of the chromosomes of recipient cells; this may bring about a genetic change in these cells. The uptake of DNA molecules is an active process requiring energy. Therefore, transformation occurs in only those species of bacteria that possess the enzymes for active transport. Further, only some cells in the cultures of these species are ***competent***, *i.e.* capable of transformation. Competent cells contain a 'competence factor', which is most likely a cell surface protein involved in binding or uptake of DNA. In a given culture, the frequency of competent cells is the maximum during the late log phase.

The mechanism of transformation is summarised below; the process may be divided into the following five steps.

1. A reversible binding of the double-stranded DNA, ***donor DNA***, to receptor sites on the surface or recipient cell.
2. An irreversible uptake of the donor DNA.
3. Degradation of one strand of the donor DNA molecules thereby making them single-stranded.

4. Integration (covalent insertion) of all or a part of the single-stranded donor DNA in the place of one strand of the homologous segment of chromosome of the recipient cell.
5. Segregation (due to semiconservative DNA replication) and phenotypic expression of the integrated donor gene or genes in the recombinant (***transformed***) cells.

Competent cells bind, take up and degrade one strand of DNA molecules from any species without any discrimination. But the integration of homologous DNA (DNA from the same species) is several orders of magnitude more frequent than that of heterologous (foreign) DNA. Very small fragments of DNA are taken up, but a minimum length of about 500 base pairs is required for their integration. The degradation of one strand of the donor DNA is random in that either of the two strands is equally likely to be degraded. During integration, the donor DNA becomes physically inserted in the place of one strand of the corresponding segment of the chromosome of recipient cell. Generally, the integrated donor DNA is about 2,000 base pairs long (about 1/200 of the bacterial chromosome). Therefore, only closely linked genes can be mapped through transformation.

Gene mapping through transformation is based on the following logic. Simultaneous transfer (from donor DNA into the recipient chromosome) of two genes (***cotransormation***) will be quite frequent if they are closely linked. The frequency of cotransformation of two genes decreases as the distance between them increases. Therefore, genes located far apart may be expected to rarely cotransform.

18.8.2. Transduction

Transfer of DNA from one bacterial strain into another by a virus and the subsequent recombination between the chromosomes of recipient cells and the introduced DNA is known as ***transduction***. The phenomenon was first described by N. Zinder and J. Lederberg in 1952. Transduction may be either (I) generalised or (2) specialized.

1. Generalized Transduction. In ***generalized transduction***, a random or nearly random segment of the bacterial chromosome is transferred by the virus. This type of transduction is mediated by some virulent phages and by certain temperate phages whose chromosomes are not integrated at specific attachment sites in the host chromosome. Some of the mature particles of these phages contain a segment of the host chromosome. These particles, called ***transducing particles***, are responsible for transduction; they are produced during the lytic cycle of these phages. In some cases, transducing particles contain only bacterial DNA, while in others they contain both bacterial and phage DNA; the bacterial DNA segment may represent 1/100 to 1/50 of the total bacterial chromosome. Since all of the genes of the host chromosome are expected to be present in a population of transducing particles, this type of transduction is called *generalized transduction.*

Bacteriophages are called virulent or temperate on the basis of their interaction with their hosts. ***Virulent phages*** always multiply and lyse the host cells, which releases several progeny phage particles. But ***temperate phages*** may either (1) multiply and lyse the host cells like virulent phages (***lytic cycle***) or (2) their chromosome may become integrated into and behave as a part of the host chromosome (***lysogenic pathway***); an integrated phage chromosome is called ***prophage***. Chromosomes of temperate phages integrate into the host chromosome at one or a few specific attachment sites. A bacterium carrying a prophage is called

lysogenic and the prophage-host relationship is known as ***lysogeny***. The genes responsible for viral multiplication and lysis of host cells (***lytic genes***) are repressed in the prophages. A lysogenic bacterial cell is immune to further infections by the same virus as it turns off the lytic genes of the new phage chromosomes as well.

Only some virulent phages mediate transduction. Of these, *E. coli* phage P1, *Salmonella* phage P22, *Bacillus subtilis* phage PBS1 and SP10 have been used extensively for the mapping of mutant sites within individual genes or genes located in a short segment of the chromosome. The host DNA introduced by a transducing particle may become integrated in the double-stranded state in place of the homologous segment of the host chromosome, or it may remain free in the cytoplasm. The former generates ***stable transduction***, while the latter gives rise to abortive transduction. In ***abortive transduction***, the nonintegrated DNA segment lying free in the cytoplasm is not degraded and is passed on to one of the two daughter cells at each cell division where genes express themselves. Cells carrying a nonintegrated DNA segment are partially diploid (***merozygotes***) and are used for complementation tests.

Transducing particles are produced at a low frequently (10^{-5} – 10^{-7}). Therefore the probability of a cell being doubly transduced (***contransduced***) for genetic markers would indicate that they are located in a short segment of the host chromosome. Further, the more closely two genes are linked, the greater would be the frequency of their cotranduction. Thus the frequency of cotransduction of three or more genes, generally three genes at a time, can be used to determine their order.

2. Specialized Transduction. Temperate phage chromosome are integrated into their host chromosomes through a process of recombination. The recombination occurs at ***specific attachment sites*** present in both phage and host chromosomes. The attachment sites of the *coliphage* (phage of *E. coli*) lambda (phage λ) chromosome (pp′) and *E. coli* chromosome (bb′) contain the following identical 15-base-pair sequence called the ***core sequence***.

5′ —— GCTTTTTTATACTAA ——→
←—— CGAAAAAATATGATT —— 5′

The site-specific recombination between *E. coli* and λ chromosomes occurs within this site, and is catalyzed by the product of *int* gene of λ in cooperation with another protein IHF (integration host factor). The λ chromosome is linear in mature phage particles. It changes into a covalently bound circular form after entering the host cell; this form may undergo integrative recombination with the host chromosome (Fig. 18.6).

Prophages undergo spontaneous excision from the host chromosomes at very low frequencies (10^{-5}/cell division). This process is generally very precise and site-specific. The free phage chromosome produced in this manner replicates and causes lysis of the host cell. The frequency of phage chromosome excision is greatly enhanced by certain treatments, *e.g.*, irradiation with ultraviolet light; this is called ***induction***. During excision, the prophage chromosome is believed to loop out forming the figure of '8' so that the two attachment sites lie next to each other. A site-specific recombination now occurs within the two attachment sites producing circular phage and host chromosomes (Fig. 18.6); in phage λ, this requires the product of phage genes *int* and *xis* plus protein IHF. This recombination is highly precise so that the excised phage chromosome is identical with the one that was originally integrated.

Occasionally, excision occurs at a site other than the original attachment site so that a portion of the phage chromosome is left in the host chromosome and a part of the latter is excised as a part of the phage chromosome. This gives rise to the ***specialized transducing particles***. Such particles carry only those host genes that are located close to the site of prophage insertion. Therefore, ***specialized transduction*** *transfers only those host genes that are closely linked with the phage attachment site*. For example, phage λ integrates between the genes *gal* (gene required for utilization of galactose) ad *bio* (gene governing the biosynthesis of biotin) of *E. coli*

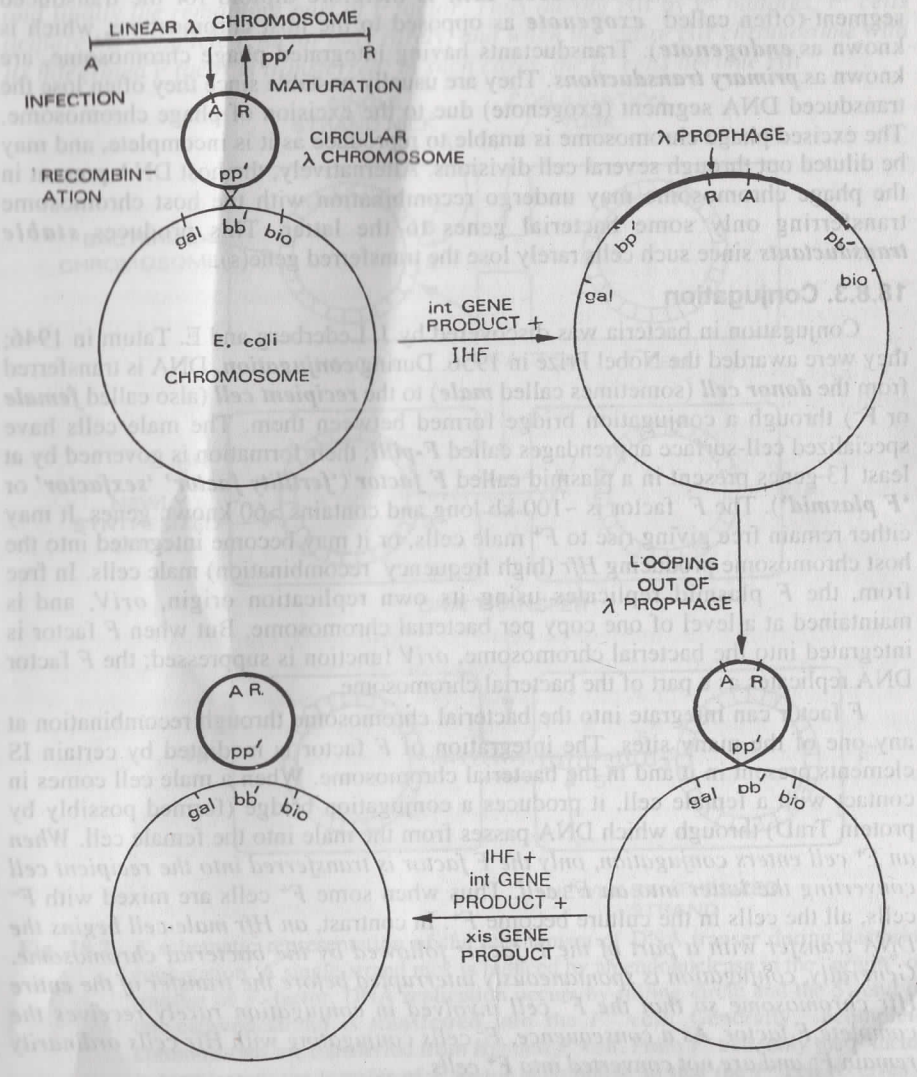


Fig. 18.6. A schematic representation of the integration into and the excision from *E. coli* chromosome of temperate phage lambda (λ) chromosome.

(Fig. 18.6), and it generally transduces only these genes. It should be noted that, of necessity, specialized transducing particles are produced at a low frequency (10^{-6} cells) by induction only; they are not produced by the natural lytic pathway of temperate phages.

When a specialized transducing particle infects a host cell, the phage chromosome (carrying a segment of the host chromosome) may integrate into the host chromosome through recombination at the attachment site. The host chromosome now contains the incomplete phage chromosome as well as the segment of host chromosome present in the transducing particle. Such a host cell, called ***transductant*** or ***transduced cell,*** is therefore diploid for the transduced segment (often called ***exogenote*** as opposed to the host chromosome, which is known as ***endogenote***). Transductants having integrated phage chromosome, are known as ***primary transductions***. They are usually unstable since they often lose the transduced DNA segment (exogenote) due to the excision of phage chromosome. The excised phage chromosome is unable to reproduce as it is incomplete, and may be diluted out through several cell divisions. Alternatively, the host DNA present in the phage chromosome may undergo recombination with the host chromosome transferring only some bacterial genes to the latter. This produces ***stable transductants*** since such cells rarely lose the transferred gene(s).

18.8.3. Conjugation

Conjugation in bacteria was discovered by J. Lederberg and E. Tatum in 1946; they were awarded the Nobel Prize in 1956. During ***conjugation***, DNA is transferred from the ***donor cell*** (sometimes called ***male***) to the ***recipient cell*** (also called ***female*** or F^-) through a conjugation bridge formed between them. The male cells have specialized cell-surface apprendages called ***F-pilli***; their formation is governed by at least 13 genes present in a plasmid called ***F factor*** (***'fertility factor' 'sexfactor'*** or ***'F plasmid'***). The *F* factor is ~100 kb long and contains >60 known genes. It may either remain free giving rise to F^+ male cells, or it may become integrated into the host chromosome producing *Hfr* (high frequency recombination) male cells. In free from, the *F* plasmid replicates using its own replication origin, *oriV*, and is maintained at a level of one copy per bacterial chromosome. But when *F* factor is integrated into the bacterial chromosome, *oriV* function is suppressed; the *F* factor DNA replicates as a part of the bacterial chromosome.

F factor can integrate into the bacterial chromosome through recombination at any one of the many sites. The integration of *F* factor is mediated by certain IS elements present in it and in the bacterial chromosome. When a male cell comes in contact with a female cell, it produces a conjugation bridge (formed possibly by protein TraD) through which DNA passes from the male into the female cell. *When an F^+ cell enters conjugation, only the F factor is transferred into the recipient cell converting the latter into an F^+ cell.* Thus when some F^+ cells are mixed with F^- cells, all the cells in the culture become F^+. In contrast, *an Hfr male cell begins the DNA transfer with a part of the F factor followed by the bacterial chromosome. Generally, conjugation is spontaneously interrupted before the transfer of the entire Hfr chromosome so that the F^- cell involved in conjugation rarely receives the complete F factor. As a consequence, F^- cells conjugating with Hfr cells ordinarily remain F^- and are not converted into F^+ cells.*

The mechanisms of transfer of both *F* factor (from F^+ cells) and *Hfr* chromosome (from *Hfr* cells) appears to be similar. Protein TraY and/or TraI nicks one strand of the *F* factor or the *Hfr* chromosome at a specific site, called the

'origin' of transfer (***oriT***) in the *F* factor. The 5′-end of nicked strand progressively separates from its complementary strand, and is transferred into the F^- cell through the conjugation tube. This separation is brought about by TraY/TraI multimer, which binds to *oriT* after nicking. The separation is very rapid (~1200 bp per second) and does not depend on DNA replication. Usually the nicked strand undergoes rolling circle replication (see, Chapter 24). The single strand transferred into the F^- cell serves as a template for the synthesis of its complementary strand, the DNA synthesis being discontinuous (Fig. 18.7). Further, the synthesis of complementary strand is simultaneous with the DNA transfer. Since the cleavage occurs within the *F* factor, the sequence of DNA transfer from *Hfr* into the F^- cells is a follows: *a part of the* F *factor located away from the* tra *region, followed by the bacterial chromosome and finally the remaining part of* F *factor. Thus during* Hfr *by* F^- *matings,* F^- *cells acquire the complete* F *factor only rarely, and when the entire* Hfr *chromosome with its integrated* F *facor is transferred; in such cases, the* F^- *cells become* Hfr.

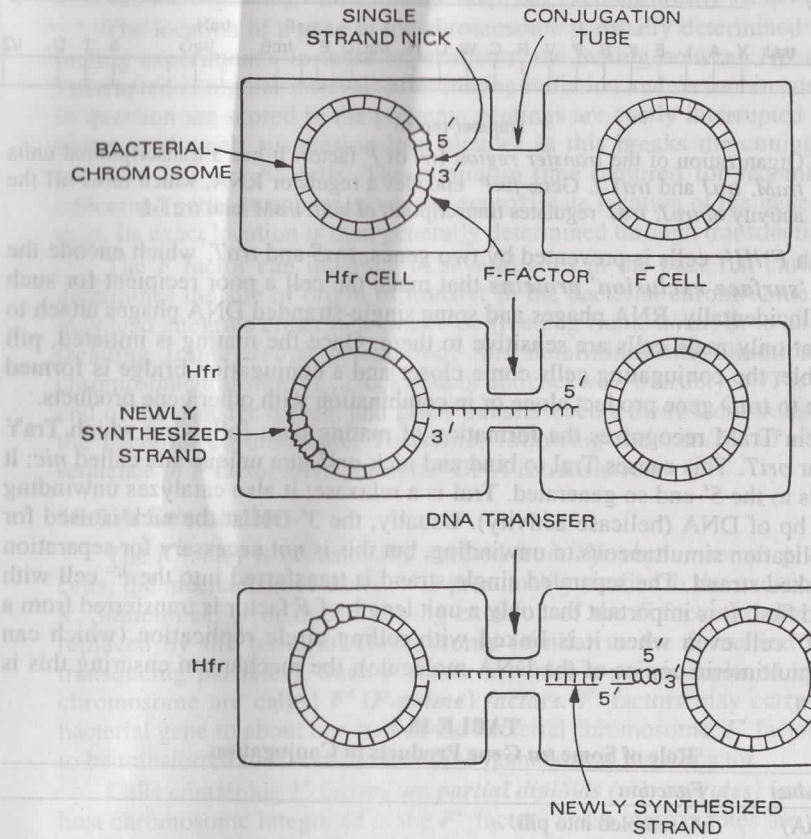


Fig. 18.7. A schematic representation of the mechanism of DNA transer during bacterial conjugation. A single-strand nick is induced by an endonuclease in the "origin" of transfer of *F* factor. DNA replication occurs by rolling circle and the 5′–end of the cleaved strand is transferred into the F^- cell. Generally, incomplete chromosomes are transferred from *Hfr* into F^- cell. From F^+ cell, only the *F* factor is transferred; the transfer of *F* factor is complete so that the F^- cell receives a complete circular *F* factor.

Conjugation is controlled by a ~33 kb region of the *F* plasmid called the ***transfer region* (*tr*)**. This region contains ~40 genes, which are named as *tra* and *trb* loci (Fig. 18.8). These genes are arranged in 3 transcriptional units: (1) *traj* encodes a regulator which turns on the other two units, (2) *traM* and (3) a 32 kb transcription unit *traY-I*. Gene *finP* is located on the opposite strand and encodes a regulator antisense RNA that turns off *traJ*; its activity requires the expression of another gene *finO*. The products of *tra* genes affect the properties of cell surface as well as are involved in DNA transfer (Table 18.3); only 4 of the genes in the *traY-I* transcription unit are concerned directly with DNA transfer.

Gene *traA* encodes the monomeric protein pilin, which is assembled into *F*-pili; the assembly is governed by at least 12 *tra* genes. *F*-pili are 2-3 μm long ~8 nm diameter cylinders with a 2 nm axial hole. A typical F^+ cell has 2-3 pili. Conjugation begins when the tip of *F*-pilus of a cell comes in contact with a F^- cell. Interaction of

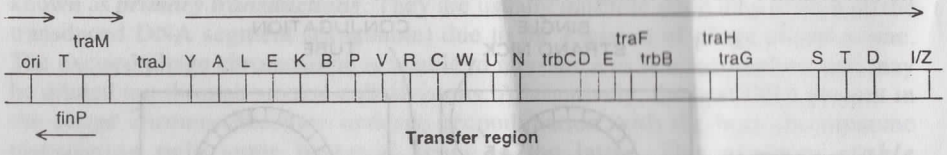


Fig. 18·8. Organization of the *transfer region* (*tr*) of *F* factor. It has 3 transcriptional units *traM, traJ* and *traY-I*. Gene *finP* encodes a regulator RNA, which turns off the activity of *traJ; traJ* regulates transcription of both *traM* and *traY-I*.

F-pili with F^+/*Hfr* cells is prevented by two genes, *traS* and *traT*, which encode the so called ***'surface exclusion' proteins*** that make the cell a poor recipient for such contacts. Incidentally, RNA phages and some single-stranded DNA phages attach to pili so that only male cells are sensitive to them. Once the mating is initiated, pili disassemble, the conjugating cells come closer and a conjugation bridge is formed either due to *traD* gene product alone or in combination with other gene products.

Protein TraM recognizes the formation of mating pair, following which TraY binds near *oriT*. This causes TraI to bind and nick *oriT* at a unique site called *nic*; it then binds to the 5′-end so generated. TraI is a relaxase; it also catalyzes unwinding of ~200 bp of DNA (helicase activity). Usually, the 3′-OH at the nick is used for DNA replication simultaneous to unwinding, but this is not necessary for separation of the nicked strand. The separated single strand is transferred into the F^- cell with the 5′-end first. It is important that only a unit length of *F* factor is transferred from a F^+ to a F^- cell even when it is linked with rolling circle replication (which can produce multimeric copies of the DNA molecule); the mechanism ensuring this is not known.

TABLE 18.3
Role of Some *tra* Gene Products in Conjugation

Gene product	*Function*
Pilin (TraA)	Assembled into pili.
TraD	May form conjugation bridge.
TraI	See TraY.
TraJ	Regulator protein; turns on expression of *traM* and *traY-I*.
TraS, TraT	Surface exclusion' proteins; prevent interaction of *F*-pili with the cell surface of F^+/*Hfr* cells.
TraY	Together with TraI nicks one strand at *oriT* and unwinds the nicked strand beginning at the 5′–end.

In E. coli, *the transfer of complete* Hfr *chromosome takes ~100 minutes, depending on the strain, the rate of transfer being constant throughout a mating.* The bacterial chromosome (***exogenote***) transferred into the F^- cell is single-stranded; it undergoes replication to become double-stranded and then may undergo recombination with the chromosome of F^- cell (***endogenote***). It can be easily seen that two (or a multiple of two) simultaneous recombinations between the exogenote and endogenote are required to transfer a segment of the former into the latter without affecting the structural integrity of the later [a single, or an odd number, of recombination(s) would generate a linear or noncircular recombinant chromosome]. Thus the segment of *Hfr* chromosome transferred during mating can be determined by the recombinants recovered from the mating. The time required, after the initiation of mating, for the transfer of any two genes is used as an estimate of the distance between them. The *E. coli* linkage map is divided into 100 intervals, each interval equating one minute time required for gene transfer during mating. The 0 interval, the beginning of the linkage map, has been arbitrarily set at *thrA* gene.

The location of a gene in the chromosome is usually determined by interrupted mating experiments. In these experiments, the mating between *Hfr* and F^- cells is interrupted at regular intervals after mating initiation and recombinants for the genes in question are scored in the progeny. Matings are easily interrupted by vigorously agitating the cell suspension in a blender as this breaks the conjugation bridges between *Hfr* and F^- cells. The minimum time required for recombinants of the concerned genes to appear, gives the approximate location of the gene in the linkage map. Its exact location is then generally determined through transduction mapping.

The *F* factor can integrate at several sites in the bacterial chromosome; this determines the site of origin of transfer in the bacterial chromosome. The *F* factor may also integrate either in a clockwise (reading frame *abcd*) or in an anticlockwise (reading frame *dcba*) orientation; this determines the direction of bacterial chromosome transfer (*i.e.*, clockwise or anticockwise transfer). *oriT* is located at one end of the transfer region, and DNA transfer occurs away from transfer region into the bacterial chromosome. Therefore, following the transfer of a short leading sequence of *F* factor DNA, bacterial DNA is transferred.

18.8.4. Sexduction

The *F* factor is occasionally excised from *Hfr* chromosomes giving rise to F^+ cells; the mechanism of excision is presumably similar to the excision of prophage λ. Sometimes, error occurs during excision of the *F* factor so that a part of it is replaced by the bacterial chromosome (similar to the formation of specialized transducing particles). Such *F* factor in which a part is replaced by the bacterial chromosome are called ***F′* (*F-prime*) *factors.*** *F′* factors may carry from a single bacterial gene to about one-half of the bacterial chromosome. *F′* factors are believed to be transferred to F^- cells in the same manner as is the *F* factor.

Cells containing *F′* factors are ***partial diploids (merozygotes)*** for the segment of host chromosome integrated in the *F′* factor. These merozygotes are unstable as the *F′* factor may be lost. Alternatively, recombination between an *F′* factor and the bacterial chromosome may lead to the integration of the *F′* factor (alongwith the bacterial genes contained in it) into the latter; such a gene transfer (mediated by *F′* factor) is called ***sexduction*** or ***F-duction***. Thus *sexduction produces partial diploidy for the segment of bacterial chromosome present in the* F′ *factor*. Therefore, sexduction provides an unique opportunity for the determination of dominance

relationships between alleles and of defining genes by complementation test. Since *F′ factors are available carrying almost any segment of* E. coli *chromosome, sexduction analysis can be carried out with almost any mutation in the bacterial genome.*

18.9. RECOMBINATION IN VIRUSES

Viral chromosomes undergo recombination during their reproduction in host cells. Therefore, a host cell must be simultaneously infected by two different strains/genotypes of a virus in order to premit their chromosomes to undergo recombination. This is achieved by infecting host cells with a mixture of the two selected virus strains; the density of virus particles is kept high so that, on an average, each host cell is infected by two or more virus particles. As a result, some host cells will be infected by particles of both the viral strains. In such cells, chromosomes of the two strains may undergo recombination to yield recombinant progeny phages. For example, a recombionation between a^+b and ab^+ chromosomes would yield the wild type a^+b^+ and the double mutant *ab* progeny.

The phenomenon of recombination in viruses was first described by A.D. Hershey and R. Rotman in 1945. Linkage maps of several phages have now been prepared and in most cases, *e.g.*, T_2, T_4 etc. copliphages, they have been found to be circular. Recombination among the r^{II} mutants (a conditional lethal) or T_4 coliphage was extensively studied by Benzer for developing the fine structure map of this locus (see, Chapter 19).

QUESTIONS

1. Describe the mechanism of transduction and transformation (or conjugation) in bacteria.
2. In *Neurospora*, gene *ag* is involved with the biosynthesis of arginine and gene *thi* with that of thiamine. Zygotes heterozygous for both *ag* and *thi* were produced by the cross *ag thi* × + +; the asci obtained from these zygotes showed the following ascospore arrangements.

Ascospore pair	*Phenotype*			
1	*ag thi*	*+thi*	*++*	*ag +*
2	*ag thi*	*+ thi*	*++*	*ag +*
3	*+ +*	*ag +*	*ag thi*	*+ thi*
4	*+ +*	*ag +*	*ag thi*	*+ thi*
Number of ascri	42	40	39	42

(+ denotes the wild type allele of the gene)
Determine if the genes *ag* and *thi* are linked or if they are segregating independently.

3. In *Neurospora*, an adenine (*ad*) and tryptophan (*trp*) requiring stock (*ad trp*) was crossed with the wild type strain (++). The asci obtained from this cross showed the following spore arrangements.

ad trp	*ad +*	*ad trp*	*ad trp*	*ad trp*	*ad +*	*ad trp*
ad trp	*ad +*	*ad +*	*+ trp*	*+ +*	*+ trp*	*+ +*
+ +	*+ trp*	*+ trp*	*ad +*	*ad trp*	*ad +*	*+ trp*
+ +	*+ trp*	*+ +*	*+ +*	*+ +*	*+ trp*	*ad +*
49	7	31	2	8	1	2

(*i*) Are the genes *ad* and *trp* linked.
(*ii*) If so, determine the frequency of recombination between them and their centromere. Draw a linkage map for them.

4. In an organism with unordered tetrads, the following data were obtained.

abe	*ab+*	*abe*	*abe*
abe	*ab+*	*a + e*	*+be*
+++	*++e*	*+b+*	*a++*
+++	*++e*	*+++*	*+++*
247	86	8	32

(*i*) Compute the frequencies of PD, NPD and T tetrads for each pair of genes and determine if the gene *a*, *b* and *e* are linked.
(*ii*) Determine the frequencies of recombination between them and prepare their linkage map.

5. Write short notes on:
(*i*) tetrad analysis,
(*ii*) generalized transduction,
(*iii*) specialized transduction,
(*iv*) recombination in viruses,
(*v*) F factor,
(*vi*) sexduction,
(*vii*) transformation,
(*viii*) conjugation,
(*ix*) recombination during mitosis.

19

Multiple Alleles

19.1. INTRODUCTION

It is generally assumed that a gene has two alternative forms called ***alleles***. Usually, one of the two alleles of a gene is dominant over the other, which is recessive. The two alleles of a gene determine the two contrasting forms of a single character, *e.g.*, round and wrinkled seed shapes in pea. But *in many cases, several alleles of a single gene are known, each governing a distinct form of the concerned trait. Such a situation is known as* ***multiple allelism***, and the many alleles of a single gene are called ***multiple alleles***. *In the case of multiple alleles, only the monohybrid ratio is observed in the* F_2 *generation.*

Many genes in both prokaryotes and eukaryotes (both plants and animals) show multiple alleles. Some examples of such genes are (1) human ABO, (2) Rh and (3) other blood groups, (4) human haemoglobin, (5) self-incompatibility in plants, (6) fur colour in rabbits and other mammals, (7) *white* eye (*w*), (8) *rosy* eye, (*ry*), (9) *lozenze* eye (*lz*) colour genes of *Drosophila,* (10) *Notched* wing (*N*) locus of *Drosophila,* (11) *waxy* (*wx*) gene of maize etc. These are only some of the genes that show multiple alleles. It is believed that most (or even all) of the genes have multiple alleles.

19.2. ABO BLOOD GROUP IN MAN

Blood groups of animals arise due to the presence, on the outer surfaces of their erythrocytes, of molecules having antigenic properties. An ***antigen*** is a molecule, which specifically interacts with an antibody. Most antigens are also immunogenic. An ***immunogen*** is a molecule, which when introduced into the system of an animal, *e.g.*, rabbit, mice, man etc., induces the production of a class of proteins, known as ***antibodies***, which specifically bind to this molecule. Most antigens are immunogenic. Ordinarily, an antigen induces the production of an antibody specific to itself. Therefore, the name of an antibody is derived by adding the word 'anti-' as a prefix to the name of the concerned antigen. When molecules of an antigen come in contact with the molecules of the antibody specific to this antigen, several antigen and antibody molecules bind together, and usually form a precipitate. this is known as ***antigen-antibody interaction.*** This interaction is highly precise and forms a very sensitive and potent diagnostic tool in many biological studies. Such diagnostic procedures are called ***serological tests.***

19.2.1. Antibodies

Antibodies belong to a class of proteins called ***immunoglobulins***. Each antibody molecule has two antigen binding sites. Antibodies are produced by *B* lymphocytes. All the antibody molecules produced by a single lymphocyte have the same antigen-

binding specificity. Every individual has a highly heterogeneous population of lymphocytes.

19.2.2. ABO Antigens

Human erythrocytes (RBCs, red blood corpuscles) contain several different types of antigens on the outer surface of their plasma membranes. These antigens give rise to the different human blood group systems. *It is of paramount significance that an animal does not ordinarily produce antibodies against those antigens, which are naturally present in its own body since it would be suicidal. Thus* ***an animal produces antibodies against only those antigens, which are foreign to its system.***

A group of antigens present on human RBCs are known as antigens A and B; they are the basis for ABO blood group. Antigens A and B are produced from antigen O or antigen H, which is produced in all individuals. Antigen O consists of a particular group that is added to proteins. Allele *i* (Section 2.3) carries a small deletion, produces an inactive enzyme and, as a result, is unable to further modify antigen O. But alleles I^A and I^B encode an active galactosyl transferase. The enzyme produced by I^A, *i.e.*, N-acetyl - galactose transferase, converts antigen O to antigen A (by adding N-galactosamine). Similarly, the enzyme galactose transferase encoded by I^B allele converts entigen *O* to antigen *B* (by adding galactose). Therefore, for simplicity, the antigen A is described as galactosamine, and the antigen B as galactose.

ABO Blood Group

- Antigen present on RBC surface.
- *Discovery*: Landsteiner in 1900.
- Groups :
  - (*i*) A: antigen A (serum, anti- B).
  - (*ii*) B: antigen B (Serum, anti-A).
  - (*iii*) AB: antigens A and B (serum, no antibody).
  - (*iv*) O: antigen O (serum, both anti-A and anti-B).
- *Genetics*: Monogenic; alleles *i*, I^A, I^B.
  - (*i*) Antigen 'O' produced in all; added to proteins.
  - (*ii*) locus *i* encodes galactosyl transferase enzyme.
  - (*iii*) I^A adds galactosomine to antigen 'O'.
  - (*iv*) I^B adds galactose to 'O'.
  - (*v*) Allele *i* encodes defective enzyme; 'O' not modified.
- Crucial in blood transfusion.

19.2.3. ABO Blood Group System

The human ABO blood group was first described by Landsteiner in 1900. Human beings are classified into four groups on the basis of the presence of A and B antigens on their RBCs. Erythrocytes of individuals belonging to ***group A*** have antigen A on their outer surface. In case of ***group B*** individuals, erythrocytes have antigen B, while RBCs of ***AB group*** individuals have both the antigens A and B. However, in many individuals, RBCs have neither antigen A nor antigen B; such individuals are placed into ***group O.*** *The blood serum of group A individuals contains antibody-B (written as anti-B for brevity), that of group B individual has anti-A, that of group AB individuals has neither anti-A nor anti-B, while the serum of group O individuals contains both anti-A and anti-B.* (***Blood serum*** is the liquid

phase in which the different types of blood cells are suspended.) ***Anti-A*** and ***anti-B*** are antibodies specific to antigens A and B, respectively.

When anti-A is introduced into a suspension of RBCs having antigen A, antigen-antibody interaction takes place causing ***agglutination*** or ***coagulation of the*** RBCs. Agglutination would also occur when anti-B is introduced into a suspension of RBCs having antigen B. ***Agglutination*** is an irreversible aggregation and association of RBCs in groups of variable sizes and is seen as clot formation.

ABO is the most important of the more than twenty blood group systems known in man. If blood transfusion is carried out at a slow rate, the antibodies present in the serum or plasma of the blood being transfused are rapidly diluted in the recipient's blood. Therefore, they are unable to react with the recipient's RBCs in sufficient quantities to cause their agglutination. On the other hand, the antigens present on the RBCs in the transfused blood are exposed to a large quantity of the antibodies present in the recipient's blood serum. Therefore, a sufficient number of antibodies would interact with the transfused RBCs causing their agglutination. Therefore, ***during blood transfusion, the antigens present in the donor (transfused) blood and the antibodies present in the recipients plasma are important.***

Thus blood from a person of A blood group (donor) cannot be transfused into a person of blood group B (recipient) (since anti-A present in the serum of recipient will interact with the antigen A present on the RBCs in the donor's blood and agglutinate them). For the same reason, blood from a group B donor cannot be transfused into a recipient of group A (Table 19.1). However, a recipient having the blood group AB can be given blood from donors of either A or B blood group (since the serum of such recipients has neither anti-A nor anti-B). Therefore, individuals with AB blood group are called ***universal recipients.*** Similarly, the RBCs of O blood group persons lack active A and B antigens. Hence their blood can be safely transfused into recipients belonging to any of the blood groups of the ABO system. Therefore, they are known as ***universal donors.***

TABLE 19.1

The Various Genotypes, their Phenotypes and the Compatible Blood Groups for Transfusion in the Case of Human ABO Blood Group System

Blood group	*Genotype*	*Antigen present on the surface of RBC's*	*Antibody present in blood serum*	*Compatible transfusion from the blood groups*
A	I^AI^A I^Ai	A (galactosamine)	Anti-B	A and O
B	I^BI^B I^Bi	B (galactose)	Anti-A	B and O
AB	I^AI^B	A and B	Nil	A, B, AB and O
O	ii	Nil	Anti-A and Anti-B	O

Notes: 1. Incompatible blood transfusions: In group A, from groups B and AB; in group B, from groups A and AB; in group AB, from none; in group O, from groups A, B and AB.

2. In the determination of compatibility for blood transfusion, Rh blood group is also important.

3. ABO blood group system is also involved in histocompatibility, *i.e.,* rejection of tissue transplants.

4. In fact, the number of blood groups in this system is more than the four listed above. *At this locus, at least five different alleles are known.*

19.2.4. Inheritance of the ABO System

The ABO blood group system is believed to be controlled by a single gene, generally designated as *I*. Gene *I* has the following three alleles: I^A, I^B and I^o or *i*. Allele I^A controls the production of antigen A, allele I^B determines antigen B, while allele *i* (I^o) does not produce an active antigen. These alleles are codominant so that individuals having the genotype I^AI^B have both the antigens A and B on their RBCs (AB blood group). Individuals with the genotype I^AI^A or I^A *i* produce antigen A (blood group A); those with the genotype I^BI^B or I^Bi produce antigen B (group B); persons with the genotype I^AI^B have have both the antigens A and B (group AB); while those having the genotype *ii* produce neither antigen A nor B (group O) (Table 19.1).

Rh Blood Group

- Detected by antibody specific to an antigen of *Macaca rhesus.*
- Due to antigen R_0 present on RBC surface.
- Rh^+: *RBCs* react with the antibody; contain active Rh_0.
- Rh^- : *RBCs* do not react with the antibody.
- Now over 30 different groups.
- *Genetic control* : multiple alleles of gene *r.*
- *Rh-incompatible marriage :*
  - (*i*) *Wife* Rh^-, hasband Rh^+.
  - (*ii*) Second and subsequent babies may suffer from haemolytic anaemia.
  - (*iii*) Prevented by injection of defective anti-Rh_0.

19.3. HUMAN Rh BLOOD GROUP SYSTEM

Human beings can be classified into two blood groups on the basis of agglutination of their RBCs by an antibody produced in response to an antigen present on the RBCs of a species of monkeys called *Macaca rhesus* (rhesus monkeys; earlier name *Rhesus macaca*). The RBCs of some persons are agglutinated by this antibody indicating that they possess the concerned antigen; such individuals are called ***Rh-positive.*** In contrast, RBCs of ***Rh-negative*** persons do not interact with this antibody, *i.e.,* they are not agglutinated, suggesting that they lack the active form of the concerned antigen. *Since this classification is based on an antigen normally found in rhesus monkeys, this blood group system was referred to as* ***rhesus blood group*** or, *more commonly,* ***Rh blood group.***

19.3.1. Rh Antigen

The antigen involved in this blood group is called Rh-antigen and is represented as Rh_0. Rh positive individuals produce active Rh_0, which is present on their RBCs. But RBCs of Rh-negative individuals do not have Rh_0 in an active state. Rh_0 is highly immunogenic, *i.e.,* induces production of antibodies specific to itself, in human beings. Therefore, *Rh blood group is also considered in the determination of blood transfusion compatibilities.*

19.3.2. Genetic Control

Initially, persons were classified only as Rh^+ (Rh-positive) or Rh^- (Rh-negative) and the genetic control of this system appeared simple. Antigen Rh was postulated

to be produced by the dominant allele *R* of a single gene (genotypes *RR* and *Rr* being Rh-positive). The recessive allele *r* does not produce an active Rh_0 so that *rr* individuals are Rh^-. But at present, several Rh-antibodies with slightly different specificity's are known and the Rh system has more than thirty different blood groups. *According to Weiner, multiple alleles of a single gene are involved in the control of this trait* (Table 19.2). Some other workers suggested the *R* to be a complex locus consisting of 3 genes *C, D* and *E*, but this hypothesis is not generally accepted (Table 19.2).

TABLE 19·2

Multiple Allelic and Pseudoallelic Interpretations of the Nine of More than 30 Different Rh Blood Groups in Humans. According to the Pseudoallelic Interpretation, Three Tightly-linked Genes (*C, D* and *E*) Control the Production of Rh_0 (Rh-antigen)

Rh blood group	*Multiple allelic symbol (after Weiner)*	*Pseudoallelic symbol (after Race, Fisher and others)*
Rh_z	R^z	*DCE*
Rh_2	R^2	*DcE*
Rh_1	R^1	*DCe*
Rh_0	R^0	*Dce*
rh_y	r^y	*dCE*
rh′	r'	*dCe*
rH″	r''	*dcE*
rh	*r*	*dce*

19.3.3. Erythroblastosis Foetalis

Rh blood group is responsible for the ***haemolytic anaemia*** of babies (***erythroblastosis foetalis***) of Rh-negative women married to Rh^+ males. In such marriages, females will be *rr*, while males will be either *Rr* or *RR*. (We have assumed only two alleles of *r*, *R* specifying an active Rh_0 and *r* specifying an inactive Rh_0 antigen) Therefore, on an average, atleast 50% of the babies borne to such couple will be Rh^+. Ordinarily, the RBCs from developing foetus do not move across the placenta into the blood stream of mother. But at the time of childbirth, RBCs from an embryo usually migrate into the blood stream of mother. If the embryo were Rh^+ and the mother were Rh^-, the Rh_0 antigen present on these RBCs will induce a strong immunological response (*i.e.*, production of anti-Rh_0) in the mother. In the subsequent pregnancies of such a mother, the RBCs of Rh^+ embryos will be exposed to the anti-Rh_0 present in the blood serum of mother. The anti-Rh_0 molecules present in the blood serum of mother freely migrate across the placenta into the blood stream of embryo. These anti-Rh_0 molecules would attach to the RBCs of Rh^+ embryo, and cause their haemolysis producing anaemia (or *erythroblastosis foetalis*) in such embryos. The intensity of haemolytic anaemia varies considerably, and in cases of severe anaemia babies may even fail to survive. Most newborns suffering from haemolytic anaemia can, however, be saved through timely and appropriate medical care, including blood transfusion.

The problem of haemolytic anaemia in Rh^+ babies from Rh incompatible marriages, i.e., *wife Rh^- and husband Rh^+, can be successfully overcome by the use of defective anti-Rh_0.* When an Rh^- mother gives birth to an Rh^+ baby, she is immediately injected with the defective anti-Rh_0. These antibodies interact with and inactivate the Rh_0 present on the RBCs that have migrated into the mother from the

Rh^+ embryo without causing agglutination of these RBCs. This prevents the production of anti-Rh_0 in the mother. Therefore, she can subsequently produce an Rh^+ baby without the risk of haemolytic anaemia to the baby.

Fur Colour in Rabbits

- Controlled by 4 alleles of gene *c*.
  - (*i*) Allele *c* produces albina (white).
  - (*ii*) Allele c^h specifies himalayan.
  - (*iii*) Allele c^{ch} specifies chinchilla.
  - (*iv*) Allele *C* gives full colour (black/gray).
- *Dominance* : $C > c^{ch} > c^h > c$.
- Gene *c* encodes tyrosinases (involved in melanin production).
- c^h allele produces temperature-sensitive (active at low temperature) tyrosinase.

19.4. FUR COLOUR IN RABBITS

The fur colour (colour of body hairs) in mammals shows considerable variation. In rabbits, the following four types of fur colour are produced by alleles of a single gene *c*: normal or full colour, albino, himalayan and chinchilla (Table 19.3). Full fur colour is produced by the dominant allele *C* or c^+ of this gene. In the presence of allele *C*, the whole body of animals is uniformly coloured (the colour may be either black, if the recessive allele of another gene *a* is in the homozygous state, or it may be agouti or gray, if the dominant allele of this gene, *A*, is present) (Chapter 13). The fur colour of chinchilla animals is lighter than 'full colour' since their fur has a mixture of both coloured and colourless hairs. In the case of himalayan animals, the main body has colourless or white fur, while the tips or extremities of their feet, ears, tail and snout are coloured. The last category of fur colour, albina, has white or colourless hairs all over the body. The allele *C* is dominant over the alleles c^h, c^{ch} and *c*; allele c^{ch} is dominant over the alleles c^h and *c*, while allele c^h is dominant over *c*.

The *allele c^h produces a temperature-sensitive form of the enzyme tyrosinase, which is involved in the production of melanin (the skin and hair pigment).* This mutant form of the enzyme is active at low temperatures but is inactive at higher temperatures. This explains why himalayan rabbits have coloured fur at the

TABLE 19.3

The Multiple Allelic Series Controlling fur Colour in Rabbits

Allele	*Genotype*	*Phenotype*
c	*cc*	Albino (no pigment; white fur colour)
c^h	$c^h c^h$, $c^h c$	Himalayan (white body trunk; only tips of ears, feet, face and tail black)
c^{ch}	$c^{ch} c^{ch}$, $c^{ch} c^h$, $c^{ch} c$	Chinchilla (fur has a mixture of coloured and white hairs over the whole body; intensity of colour less than that in full coloured).
c^+ or *C*	$c^+ c^+$, $c^+ c^{ch}$, $c^+ c^h$, $c^+ c$	Full coloured (entire body coloured black or gray).

Dominance relationship* = $c^+ > c^{ch} > c^h > c$

*$C^+ > c^{ch}$ etc., allele c^+ dominant over allele c^{ch} etc.

extremities of their bodies, Since these parts become cooler than the rest of their bodies, the temperature-sensitive tyrosinase is able to function normally and produce melanin in these areas.

Multiple Alleles: Characteristics

- Map at the same locus.
- Wild type allele generally dominant.
- Do not ordinarily show complementation.
- Show monohybrid ratios in F_2.
- No recombination; this was the classical concept; *recombination does occur between true alleles.*
  Complementation: Heterozygote for two mutant alleles, *e.g.*, m_1, m_2, shows wild type phenotype.

19.5. CHARACTERISTICS OF MULTIPLE ALLELES

The following features have been suggested as indicative of a true multiple allelic series (***the classical concept of multiple alleles***).

1. *Multiple alleles map at the same locus* (***locus*** = a definite position in the chromosome).
2. *There is no crossing over between the members of a multiple allelic series.* Therefore, none of the progeny from individuals heterozygous for any two mutant alleles of a multiple allelic series have the wild type allele of the gene. (*But it is now well known that recombination between multiple alleles does take place.*)
3. *The wild type (normal) allele of a multiple allelic series is almost always dominant over the alleles producing mutant phenotypes.* The mutant alleles may show partial or complete dominance among themselves.
4. When a strain homozygous for a mutant allele of a multiple allelic series is crossed with another strain homozygous for another mutant allele belonging to the same series, the F_1 individuals show the mutant phenotype only; *they do not show the wild type phenotype.* In other words, ***multiple alleles do not show complementation.***
5. Further, F_2 *generations from such crosses show typical monohybrid ratio for the concerned trait.*

Defining A Gene

- *Functional definitions* :
  (*i*) Governs a trait (classical).
  (*ii*) Encodes an enzyme (Beadle and Tatum, 1941).
  (*iii*) Specifies a polypeptride (***universal view***).
- *Operational* :
  (*i*) Lack of recombination (classical, no more valid).
  (*ii*) *Lack of complementation* (usually).
  The above defines a *structural gene.*
- *Element* : DNA sequence to which regulatory proteins bind, *e.g.*, promoter.

19.6. DELINEATION OF A GENE

The decision whether members of a multiple allelic series are located in a single gene or in two or more separate but closely-linked genes is generally based on the complementation test. When ***heterozygotes for two mutant alleles* (affecting the same trait) *have the mutant phenotype, they are placed in the same gene. But if they have the wild type phenotype, the mutant alleles are considered to be located in two different genes.*** When the F_1 from two mutant strains has the normal/wild type (= nonmutant) phenotype, it is called ***complementation*** (Fig. 19.1). Complementation is used as a basis for placing mutant alleles in the same or different genes. This is called ***complementation test.*** The results from complementation tests are precise, highly reliable and they permit an operational demarcation of genes. As a rule, ***mutant alleles located in the same gene do not show complementation, while those located in different genes show complementation.*** (Fig. 19.1). Benzer proposed the term ***cistron*** (derived from *cis-trans*) for such a functional unit. However, ***most geneticists now use the term gene in the same sense as cistron.***

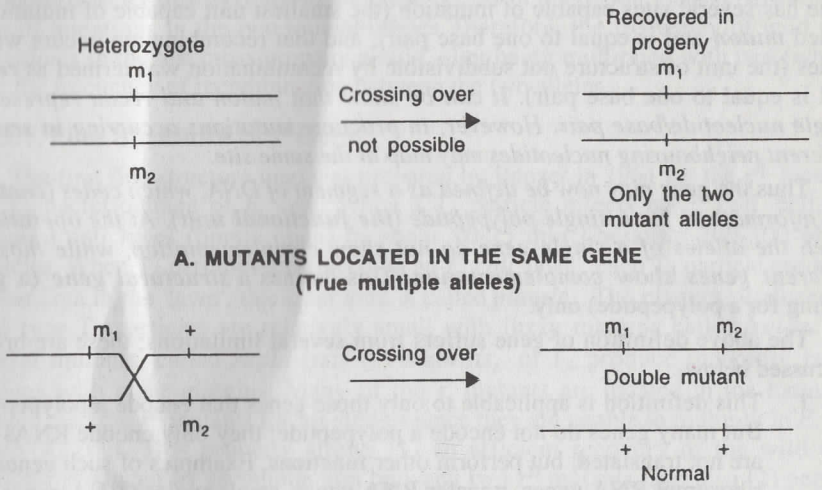


Fig. 19.1. Classification of multiple alleles as 'true multiple alleles' and as 'pseudoalleles' on the basis of recombination between them. This classification is based on the classical concept that crossing over occurs between two different genes only; *this concept is no more valid since intrangenic recombination does occur.*

19.7. COMPLEX LOCI AND PSEUDOALLELES

In case of many loci of eukaryotes, mutant alleles affecting the same function show recombination, and some of these alleles also show complementation. Such loci are termed as ***complex loci*** since they are suspected to be composed of two or more different genes affecting the same function. The multiple alleles observed at a complex locus are called ***pseudoalleles*** since all of them are not strictly allelic; they are presumed to be located in different genes of the complex locus. *The question whether a suspected complex locus is really composed of two or more separate*

genes or not can be finally resolved only when the complete nucleotide sequence of the concerned locus becomes available.

19.8. THE CHANGING CONCEPT OF 'GENE'

According to the ***classical concept***, gene is the unit of function, recombination and mutation. Thus a gene is considered to control the inheritance of one character (***unit of function***), to be indivisible by recombination and to be the smallest unit capable of mutation.

But it was soon realized that a gene does not produce a character by itself, although it may exercise the major control on its development. During 1940's, it became evident that a gene controlled a single biochemical reaction by directing the production of a single enzyme (***one-gene-one-enzyme hypothesis***). But a little later, it was shown that one gene produces a single polypeptide and not one enzyme as many enzymes (and other proteins) have more than one polypeptide in their molecules (*e.g.*, haemoglobin has 4 polypeptides) (***one-gene-one polypeptide hypothesis***). It was further established that DNA is the genetic material, that a single gene has several sites capable of mutation (the smallest unit capable of mutation is called ***muton*** and is equal to one base pair), and that recombination occurs within genes (the unit of structure not subdivisible by recombination was termed as ***recon*** and is equal to one base pair). *It can be show that muton and recon represent a single nucleotide/base pair. However, in practice, mutations occurring in several different neighbouring nucleotides may map at the same site.*

Thus the *gene may now be defined as a segment of DNA, which codes (contains the information) for a single polypeptide* (***the functional unit***). *At the operational level, the alleles of a single gene do not show complementation, while those of different genes show complementation.* This defines a ***structural gene*** (a gene coding for a polypeptide) only.

The above definition of gene suffers from several limitations; these are briefly discussed below.

1. This definition is applicable to only those genes that encode a polypeptide. But many genes do not encode a polypeptide; they only encode RNAs that are not translated, but perform other functions. Examples of such genes are ribosomal RNA genes, transfer RNA genes, small nuclear RNA genes etc. The above definition, obviously, does not include such genes. Genes that encode RNA or protein are called ***structural genes***; such genes yield a diffusible product.
2. Many genes, on the other hand, do not give rise to a diffusible product, They, however, serve as binding sites for regulatory and other DNA binding proteins, *e.g.*, promoter, operator etc. These genes are called ***elements.*** Although elements do not gives rise to an RNA/protein product, they are essential for normal function of structural genes and the genomes.
3. Further, many genes do show intragenic complementation, *i.e.*, alleles of the same gene show complementation. In such cases, complementation alone is unable to clearly identify the alleles of a given gene.

19.9. FINE STRUCTURE OF GENES

The *fine structure of a gene* is essentially the linkage map of multiple alleles of a gene, and it depicts the location of these alleles within the gene. The procedure used

Fine Structure of Genes

- Linkage map of multiple alleles.
- First prepared by Benzer (r^{II} locus of T4 phage of *E. coli*).
- Prepared for many genes of maize etc.
- Benzer proposed :
  - (*i*) *Cistron*: unit of function; lack of complementation between mutants located in one cistron;
  - (*ii*) *Recon*: Unit of recombination,
  - (*iii*) *Muton: unit* of mutation. (both represent one base-pair).

for this purpose is basically the same as that used for preparation of linkage maps of different genes. In eukaryotes, two strains homozygous for different alleles of a multiple allelic series are crossed, and the resulting F_1 is test-crossed to one of the mutants. In the test cross progeny, the frequency of individuals having the wild type phenotype for the concerned trait is determined. The wild type allele will be produced if a crossing over within the gene occurs in the region located between the sites of mutations in the two mutant alleles Therefore, the frequency of wild type individuals in the test cross progeny would, after being multiplied with two (why?), give the frequency of recombination between the two alleles.

19.9.1. Fine Structure of r^{II} Gene

The first fine structure map was prepared by Benzer in 1961 for the r^{II} locus of the T_4 bacteriophage of *E. coli*. When a T_4 particle infects an *E. coli* cell, the bacterial cell lyses (bursts open) in about 20-25 min liberating 200-300 phage particles. When T_4 particles are placed on an *E. coli* 'lawn', each T_4 particle produces a clear area in the 'lawn'; this clear area is called ***plaque***. The plaques produced by wild type T_4 particles are relatively small with fuzzy margins called ***halos***. But several mutants, called ***rapid lysis* (r) *mutants,*** of T_4 produce markedly larger plaques with clear margins. Many of the *r* mutants are located in the locus r^{II}. Benzer used the r^{II} mutants for his studies since they are unable to multiply in *E. coli* strain K12(λ). This property allows (1) an efficient recovery of the wild type particles produced by recombination between two r^{II} mutants, and also (2) permits an easy detection of complementation between two r^{II} mutants.

19.9.1.1. Complementation Test

Benzer isolated 3,000 r^{II} mutants of T_4 and used them for complementation test. This was done by infecting *E. coli* K12(λ) cells with a mixture of two r^{II} mutants at a time. *If the mutants were able to multiply* [in *E. coli* K12(λ)], it showed complementation, while a failure to do so indicated lack of complementation. Using this approach, Benzer classified the r^{II} mutants into two arbitratry groups called *A* and *B*. *Mutants from group A showed complementation with those belonging to group B, but mutants belonging to the same group did not show complementation.* Benzer proposed that the r^{II} locus consists of two cistrons, *A* and *B* (Fig. 19.2). The r^{II} mutants of group *A* were located in cistron *A*, while those of group *B* were situated in cistron *B*.

Thus ***cistron*** may be defined as an operational unit of inheritance detected by complementation test; mutants located in the same cistron do not show complementation.

genes or not can be finally resolved only when the complete nucleotide sequence of the concerned locus becomes available.

19.8. THE CHANGING CONCEPT OF 'GENE'

According to the ***classical concept***, gene is the unit of function, recombination and mutation. Thus a gene is considered to control the inheritance of one character (***unit of function***), to be indivisible by recombination and to be the smallest unit capable of mutation.

But it was soon realized that a gene does not produce a character by itself, although it may exercise the major control on its development. During 1940's, it became evident that a gene controlled a single biochemical reaction by directing the production of a single enzyme (***one-gene-one-enzyme hypothesis***). But a little later, it was shown that one gene produces a single polypeptide and not one enzyme as many enzymes (and other proteins) have more than one polypeptide in their molecules (*e.g.*, haemoglobin has 4 polypeptides) (***one-gene-one polypeptide hypothesis***). It was further established that DNA is the genetic material, that a single gene has several sites capable of mutation (the smallest unit capable of mutation is called ***muton*** and is equal to one base pair), and that recombination occurs within genes (the unit of structure not subdivisible by recombination was termed as ***recon*** and is equal to one base pair). *It can be show that muton and recon represent a single nucleotide/base pair. However, in practice, mutations occurring in several different neighbouring nucleotides may map at the same site.*

Thus the *gene may now be defined as a segment of DNA, which codes (contains the information) for a single polypeptide (**the functional unit**). At the operational level, the alleles of a single gene do not show complementation, while those of different genes show complementation.* This defines a ***structural gene*** (a gene coding for a polypeptide) only.

The above definition of gene suffers from several limitations; these are briefly discussed below.

1. This definition is applicable to only those genes that encode a polypeptide. But many genes do not encode a polypeptide; they only encode RNAs that are not translated, but perform other functions. Examples of such genes are ribosomal RNA genes, transfer RNA genes, small nuclear RNA genes etc. The above definition, obviously, does not include such genes. Genes that encode RNA or protein are called ***structural genes***; such genes yield a diffusible product.
2. Many genes, on the other hand, do not give rise to a diffusible product, They, however, serve as binding sites for regulatory and other DNA binding proteins, *e.g.*, promoter, operator etc. These genes are called ***elements***. Although elements do not gives rise to an RNA/protein product, they are essential for normal function of structural genes and the genomes.
3. Further, many genes do show intragenic complementation, *i.e.*, alleles of the same gene show complementation. In such cases, complementation alone is unable to clearly identify the alleles of a given gene.

19.9. FINE STRUCTURE OF GENES

The *fine structure of a gene* is essentially the linkage map of multiple alleles of a gene, and it depicts the location of these alleles within the gene. The procedure used

Fine Structure of Genes

- Linkage map of multiple alleles.
- First prepared by Benzer (r^{II} locus of T4 phage of *E. coli*).
- Prepared for many genes of maize etc.
- Benzer proposed :
  - (*i*) *Cistron*: unit of function; lack of complementation between mutants located in one cistron;
  - (*ii*) *Recon*: Unit of recombination,
  - (*iii*) *Muton: unit* of mutation. (both represent one base-pair).

for this purpose is basically the same as that used for preparation of linkage maps of different genes. In eukaryotes, two strains homozygous for different alleles of a multiple allelic series are crossed, and the resulting F_1 is test-crossed to one of the mutants. In the test cross progeny, the frequency of individuals having the wild type phenotype for the concerned trait is determined. The wild type allele will be produced if a crossing over within the gene occurs in the region located between the sites of mutations in the two mutant alleles Therefore, the frequency of wild type individuals in the test cross progeny would, after being multiplied with two (why?), give the frequency of recombination between the two alleles.

19.9.1. Fine Structure of r^{II} Gene

The first fine structure map was prepared by Benzer in 1961 for the r^{II} locus of the T_4 bacteriophage of *E. coli*. When a T_4 particle infects an *E. coli* cell, the bacterial cell lyses (bursts open) in about 20-25 min liberating 200-300 phage particles. When T_4 particles are placed on an *E. coli* 'lawn', each T_4 particle produces a clear area in the 'lawn'; this clear area is called ***plaque***. The plaques produced by wild type T_4 particles are relatively small with fuzzy margins called ***halos***. But several mutants, called ***rapid lysis* (r) *mutants,*** of T_4 produce markedly larger plaques with clear margins. Many of the *r* mutants are located in the locus r^{II}. Benzer used the r^{II} mutants for his studies since they are unable to multiply in *E. coli* strain K12(λ). This property allows (1) an efficient recovery of the wild type particles produced by recombination between two r^{II} mutants, and also (2) permits an easy detection of complementation between two r^{II} mutants.

19.9.1.1. Complementation Test

Benzer isolated 3,000 r^{II} mutants of T_4 and used them for complementation test. This was done by infecting *E. coli* K12(λ) cells with a mixture of two r^{II} mutants at a time. *If the mutants were able to multiply* [in *E. coli* K12(λ)], it showed complementation, while a failure to do so indicated lack of complementation. Using this approach, Benzer classified the r^{II} mutants into two arbitratry groups called *A* and *B*. *Mutants from group A showed complementation with those belonging to group B, but mutants belonging to the same group did not show complementation.* Benzer proposed that the r^{II} locus consists of two cistrons, *A* and *B* (Fig. 19.2). The r^{II} mutants of group *A* were located in cistron *A*, while those of group *B* were situated in cistron *B*.

Thus ***cistron*** may be defined as an operational unit of inheritance detected by complementation test; mutants located in the same cistron do not show complementation.

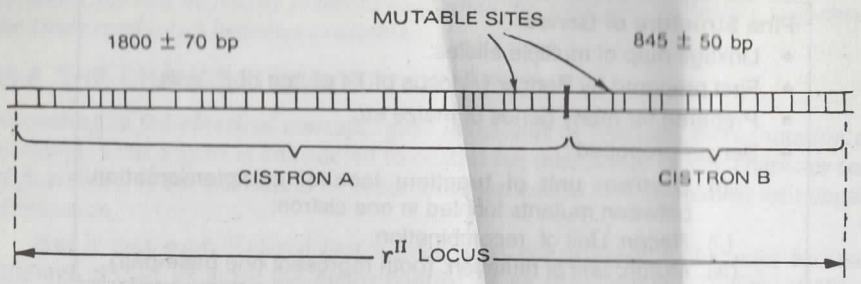


Fig. 19.2. A fine genetic map of the r^{II} locus of T_4 phage of *E. coli;* some of the more than 300 mutable sites of the locus are shown by vertical bars.

19.9.1.2. Recombination Mapping

In the next step, Benzer studied recombination between mutants located in the same cistron. To allow recombination, he infected *E. coli* strain *B* cells with a mixture of 2 different r^{II} mutants located in the same cistron (r^{II} mutants multiply in strain *B* cells). The resulting lysate was them plated on a lawn of *E. coli* K12(λ) to detect the presence of the wild type allele of r^{II}. (Wild type allele will be produced by recombination between the two r^{II} mutants.) In this manner, recombination frequencies among the different r^{II} mutants were estimated and the fine genetic map of r^{II} locus was prepared (Fig. 19.2).

The linkage mapping of the 3,000 r^{II} mutants was a gigantic task; this was made a lot easier and faster by the ingenious technique of deletion mapping. In ***deletion mapping***, a set of deletion mutants is first used for paired infection with a new or unknown mutant. In this set, each mutant carries a deletion for a different region of the concerned gene, so that they together represent the whole region of the gene. When a deletion r^{II} mutant is used for mixed infection with another r^{II} mutant, *wild type phage particles will be recovered only if the new mutant is located in the region that is intact in the deletion mutant.* Therefore, if wild type particles are not recovered after mixed affection, it becomes obvious that the new mutant is located within the region missing in the deletion mutant. Therefore, the new mutant now needs to be analyzed for recombination frequency with only those r^{II} mutants that are known to be located in the concerned region; this reduces drastically the task of mapping.

19.9.1.3. General Conclusions

The fine structure of r^{II} locus allows the following general conclusions.

1. The various r^{II} mutants mapped at about 400 different sites in cistrons *A* and *B*. Benzer proposed the terms ***recon*** and ***muton*** to specify units of recombination and mutation, respectively. Both muton and recon represent one base pair.
2. The frequency of mutation at different sites was quite variable.
3. Some sites showed very high frequency of mutation; these are called ***hot spots.***

19.9.2. Fine Structure of Genes in Eukaryotes

Complementation and recombination tests have been used to prepare fine structure maps of several eukaryotic genes. One of the best examples of such a gene

is the *rosy* (*ry*) eye locus of *Drosophila*. The different alleles of *ry* locus map at 10 different sites on the basis of recombination frequency (Fig. 19.3). Many of the *ry* mutants do not show complementation with each other (shown above the line in Fig. 19.3), while several others (depicted below the line) show complementation.

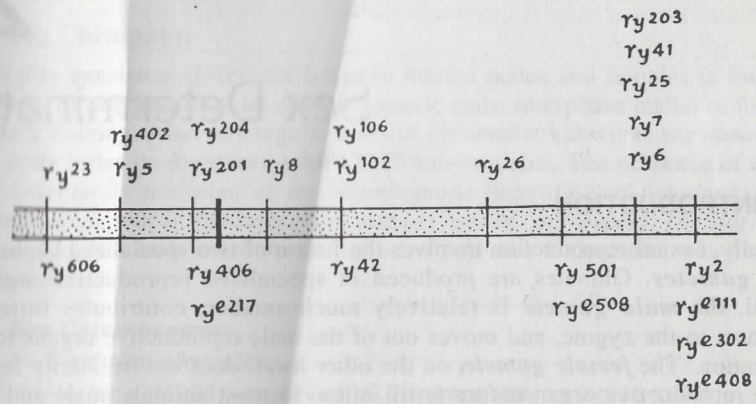


Fig. 19.3. Fine structure of the *rosy* locus of *Drosophila melanogaster*. This gene codes for xanthine dehydrogenase. Alleles shown above the line do not show complementation, while those depicted below the line exhibit complementation. The mutants denoted by the prefix *e* to their superscripts produce active enzymes with altered electrophoretic mobilities.

Similar fine structure maps have been prepared for many other eukaryotic genes, *e.g., white (w)* eye, *notch* (*N*) wing, *lozenge* (*lz*) eye, *zetse* (another eye colour locus near the white locus) etc., loci of *Drosophila, waxy (wx)* and other loci of maize, some loci of yeast and genes of some other animals and plants.

19.9.3. The Ultimate Fine Structure of Genes

The ultimate fine structure of a gene is its complete nucleotide or base sequence. Complete base sequences of many genes, of both prokaryotes and eukaryotes, have been determined. An ambitious genome project aims to ultimately determine the complete base sequence of the human genome. Several countries are participating in this mammoth project and rapid progress is being made. In addition, researches are underway to determine the complete base sequence of the genomes of many other organisms (Chapter 26).

QUESTIONS

1. Explain the phenomenon of multiple alleles with the help of an appropriate example.
2. Discuss the ABO and Rh blood groups in humans and explain their genetic basis.
3. Write short notes on the following:
   (*i*) antibodies,
   (*ii*) antigen-antibody interaction,
   (*iii*) fur colour in rabbits,
   (*iv*) complementation test,
   (*v*) pseudoalleles,
   (*vi*) the changing concept of gene.

20

Sex Determination

20.1. INTRODUCTION

Generally, sexual reproduction involves the fusion of two specialized haploid cells called ***gametes***. Gametes are produced in specialized reproductive organs. In general, the ***male gamete*** is relatively much smaller, contributes little or no cytoplasm to the zygote, and moves out of the male reproductive organs to effect fertilization. The ***female gamete,*** on the other hand, does not ordinarily leave the female reproductive organ before fertilization. In most animals, male and female reproductive organs are found in different individuals, which are referred to as ***male*** and ***female***. But in most plants, male and female reproductive organs are found in the same flower (***bisexual*** or ***hermaphrodite***) or in different flowers of the same plant (***monoecious***, *e.g.,* maize, castor, coconut etc.). In some animals also (*e.g.,* earthworm, nematodes etc.), male and female reproductive organs are found in the same individual. But in some crops (*e.g.,* papaya, *Asparagus* etc.) and several wild plant species, male and female flowers are produced on different plants (***dioecious***). Thus most animals and several plant species consist of male and female individuals that differ from each other in respect of their reproductive organs.

Sex Characters

- Characters that differ consistently between males and females of a species.
- *Primary* : concern gonads (gamete producing organs).
- *Secondary* : other sex characters including genitalia.
- *Barr body* : A dark staining chromatin mass present in interphase nucleus (somatic cells) of human females; heterochromatinized *X* ; basis of sex test.

20.2. SEX CHARACTERS

Ordinarily, male and female flower-bearing plants of a species cannot be differentiated from each other, except on the basis of the flowers they produce. But in most animals, male and female individuals differ for many characteristic features; these characters are called ***sex characters.***

20.2.1. Primary and Secondary Sex Characters

Sex characters may be grouped into two categories: (*i*) primary and (*ii*) secondary sex characters. ***Primary sex characters*** are related to the gonads or the gamete producing organs of male and female individuals. ***Secondary sex characters*** include all characters other than gonads that show consistent differences between

male and female individuals of a species. These include, genital ducts, genitalia, many body characteristics that include, in humans, facial hairs, mammary gland development, pitch of voice, degree of muscle development and subcutaneous fat deposition etc.

20.2.2. Sex Chromatin

A highly consistent difference between human males and females is found in the cytology of interphase nuclei of their somatic cells. Interphase nuclei of females have a dark staining, relatively regular mass of chromatin, known as ***sex chromatin*** or ***Barr body*** (after its discover), which is absent in males. The presence of a Barr body *is universally accepted as an unambiguous indication of femaleness of a human individual, and is the basis of* ***sex test*** *in cases of controversy. The Barr body is produced by heterochromatinization of one of the two X chromosomes found in human females.*

Sex Chromosomes

- Synonym: *allosomes.*
- Henking (1891) observed *X* chromosome in some male insects.
- Wilson and Stevens (1905) gave the name *X* chromosome. Showed its presence in both males and females.
- Stevens (1908) discovered *Y* chromosome in *Drosophila.*
- Wilson (1909) explained the role of *X* chromosome in sex determination.
- Present differentially in males and females.
  - (*i*) *Y* chromosome confined to either males (man) or females (birds).
  - (*ii*) *X* chromosome present in one copy in one sex but in two copies in the other.
- Play a role in sex determination.
- *Autosomes* : identical in males and females.

20.3. HISTORICAL

The differences in sex characters of male and female animals are perhaps the most conspicuous differences between individuals of a species. Sex is also perhaps the most important single factor influencing the usefulness of many domestic animals, *e.g.,* cattle, poultry etc. Consequently, man began attempts very early in the history to seek answer to the baffling question as to why some new-borns were male and others female ?

The first experimental evidence that ultimately led to the discovery of sex chromosomes was the observation of *X* chromosomes in some male insects by Henking in 1891. He named it as *X*-body but could not suggest its significance. A few years later, in 1905, Wilson and Stevens demonstrated that both male and female *Protenor* had the supernumerary chromosome; males had one, while females had two of these chromosomes. They called this supernumerary chromosomes *X* chromosome. The *Y* chromosome was discovered and so named by Stevens in 1908 in *Drosophila*. In 1909, Wilson proposed the involvement of *X* chromosome in sex determination in these insects.

20.4. MECHANISMS OF SEX DETERMINATION

The mechanisms of sex determination are essentially similar in both plants and animals. But for convenience, the mechanisms operating in these two groups of organisms are described separately.

Mechanism of Sex Determination: Animals

- *Environmental :*
  - (*i*) Association with females; *e.g., Bonelia.*
  - (*ii*) Egg size; *e.g., Dinophilus.*
  - (*iii*) Egg incubation temperature; *e.g.,* turtles, alligators, crocodiles etc.
- *Chromosomal* :
  - (*i*) *XX* Female, *XO* male (grasshopper, *Protenor,* other Orthoptera).
  - (*ii*) *XO* Female, *XX* male (some moths, *e.g., Fumea*).
  - (*iii*) *XX* Female, *XY* male (most common; man, mouse, *Drosophila* etc.)
  - (*iv*) *XY* Female, *XX* male (birds, reptiles, some insects like silkworm etc.).
  - (*v*) Diploid (2*n*) female, haploid (*n*) male (hymenopteran insects, *e.g.,* honey bees).
- *Genic* : Single gene determines sex; autosomal mutations in man, *Drosophila* etc.,

20.5. SEX DETERMINATION IN ANIMALS

The various mechanisms of sex determination may be classified into three main groups: (1) environmental, (2) chromosomal, and (3) genic.

20.5.1. Environmental Sex Determination

In several animals, sexual differentiation is determined by environmental factors. In different animals, different environmental factors play this role; some important factors are (1) association with families, (2) egg size and (3) incubation temperature.

20.5.1.1. Association with Females

The larvae of the sea worm *Bonelia* are sexually undifferentiated. *Those larvae that attach to the proboscis of female worms develop into males.* In contrast, those *larvae that do not attach to female worms and remain free-living develop into females* (Fig. 20.1).

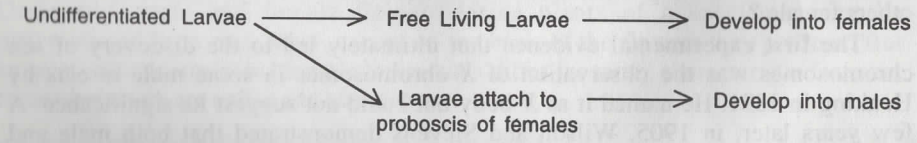


Fig. 20.1. Sex differentiation in *Bonelia* based on the mode of living of sexually undifferentiated larvae.

20.5.1.2. Egg Size

In the case of another sea worm, *Dinophilus*, the size of eggs appears to determine the sex of animals developing from them. Animals developed from eggs

of relatively larger size are females, while those obtained from comparatively smaller eggs are males.

20.5.1.3. Incubation Temperature

In some animals like turtles, alligators and crocodiles etc., the temperature of egg incubation has a critical effect on the sex of developing embryos. In some cases, (1) high (30-35°C) incubation temperatures yield only females, while low (23-28°C) temperatures produce only males; intermediate temperature (28-30°C) give both males and females. (2) In some others, high (>33°C) temperatures give rise to only males, while low (< 30°C) temperatures produce only females; both males and females are obtained at intermediate temperatures. (3) In still some others only females are produced at both high (>30°C) and low (<25°C) temperatures.

X and Y Chromosomes

- Contain a homologous region (human X^s homologous to Y^L); genes in this region show normal disomic inheritance.
- *X* and *Y* pair in this region; this ensures their regular separation.
- Only genes in X^L show sex - linkage.
- Human *Y* contains 132 loci (upto 1994) , including *SRY.*

20.5.2. Chromosomal Sex Determination

In a vast majority of animals, male and female individuals ordinarily differ from each other in respect of either the number or the morphology of the homologues of one chromosome pair; this chromosome is referred to as ***sex chromosome*** or ***allosome***. On the other hand, those chromosomes whose number and morphology do not differ between males and females of a species are called ***autosomes***. There are two types of sex chromosome, *X* and *Y*. ***X chromosome*** *is found in both males and females, although one sex has only one, while the other sex has two X chromosomes. In contrast,* ***Y chromosome*** *ordinarily occurs only in one of the two sexes of a species,* e.g., *male mice,* Drosophila, *humans etc., and female birds, reptiles etc.* In addition, *Y chromosome contains mostly heterochromatin* and was (upto 1959) generally regarded as inactive. But by 1994, 132 loci had been mapped (by deletion mapping) on human *Y*, and *SRY* gene (encoding *TDF*) was cloned. The *X* chromosome is composed of euchromatin, and many of the genes located in it show a peculiar pattern of inheritance called sex linkage (Chapter 21).

The different mechanisms of chromosomal sex determination may be grouped into five classes: (1) *XX* female, *XO* male; (2) *XO* female, *XX* male; (3) *XX* female, *XY* male; (4) *XX* male, *XY* female; and (5) diploid (2*n*) female, haploid (*n*) male.

20.5.2.1. XX Female, XO Male

In grasshoppers, *Protenor* and many other insects belonging to Orthoptera, *females have two X chromosomes, while males have only one X. Consequently, somatic cells of females have one chromosome more than those of males.* During oogenesis, the two *X* chromosomes in females pair regularly and at AI they separate and pass to the opposite poles. Thus *each of the four eggs produced by one meiotic event receives one X chromosome.* Clearly, all the eggs produced by females are alike with respect to the *X* chromosome; hence females are known as ***homogametic sex,*** *i.e.,* the sex producing only one type of gametes with respect to the sex

chromosomes. In the males, the single *X* chromosome remains unpaired during prophase I and MI; generally, it passes to one of the two poles at AI. As a result, two of the four sperms produced by one meiotic event receive one *X* chromosome, while the remaining two sperms do not receive any *X* chromosome. Thus *half of the sperms produced by males have one X chromosome, while the other half have none; therefore, males are known as heterogametic sex.*

The union of a sperm having an X chromosome with any egg produces a zygote having two X chromosomes (XX); such zygotes would develop into female individuals. But when a sperm without an X chromosome fertilizes an egg, an XO zygote is obtained; such individuals would develop into males (Fig. 20.2). *Thus one half of the progeny from each mating are females, while the other half are males.*

20.5.2.2. XO Female, XX Male

This system of sex determination is known in some insect species only, *e.g., Fumea.* In such species, females have only one *X* chromosome. As a result, they produce two types of eggs, half having an *X* chromosome and the other half having none. The males, on the other hand, have two *X* chromosomes; all the sperms, therefore, have one *X* chromosome. Thus in *Fumea* and other such species, females are the heterogametic sex, and males are the homogametic sex. The union of a sperm with an *X* chromosome containing egg produces an *XX* zygote, such zygotes develop into males. But fertilization of an egg without an *X* chromosome by any sperm gives rise to an *XO* zygote; such zygotes develop into females (Fig. 20.2). This system is essentially the reverse of that operating in grasshoppers.

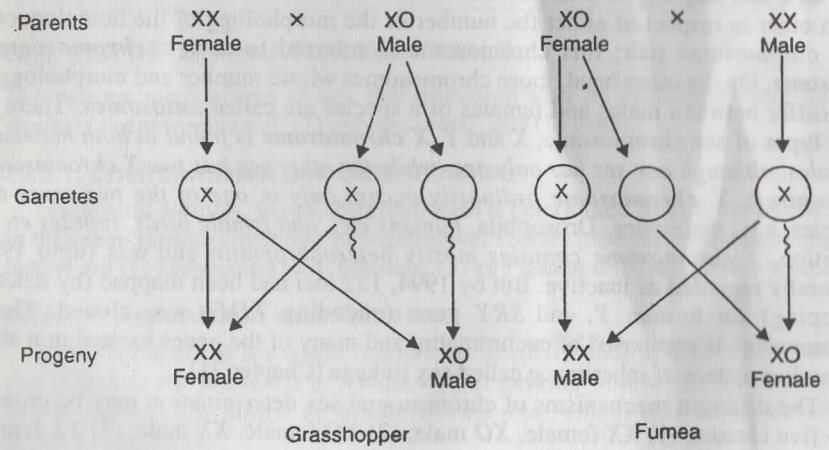


Fig. 20.2. Chromosomal basis of sex determination in grasshopper and *Fumea*. In grasshopper, the male produces two types of sperms, *X* and *O* (heterogametic sex), while in *Fumea* the female is the heterogametic sex. *O signifies the absence of an X chromosome.*

20.5.2.3. XX Female, XY Male

This system is the most common among animals. It is found in humans, mice, most other mammals, Hemiptera, Coleopteran, Diptera (*e.g., Drosophila*, house fly etc.), some fishes, some amphibia etc. In these species, females are the homogametic sex as they have two *X* chromosomes. Therefore, all the eggs produced by females

have one *X* chromosome each. The males, on the other hand, have one *X* and one *Y* chromosome. Therefore, half of the sperms produced by them have an *X* chromosome, and the rest have a *Y* chromosome (Fig. 20.3). Thus the males of these species are the heterogametic sex.

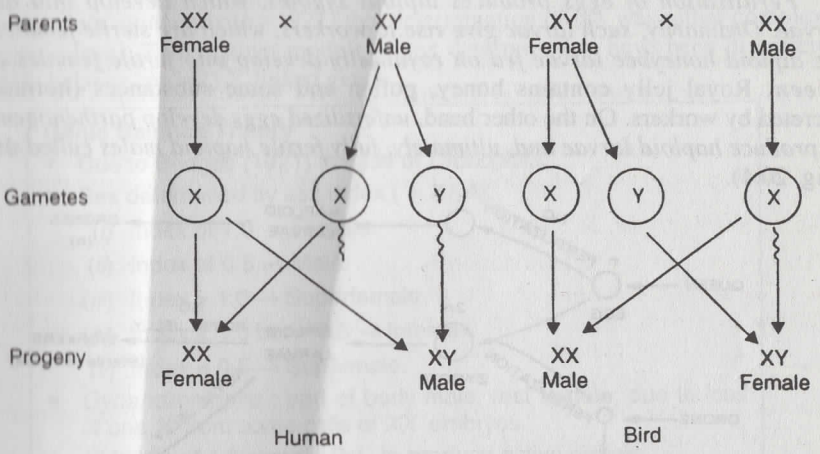


Fig. 20.3. Chromosomal basis of sex determination in man and birds. In man, male (*XY*) is the heterogametic sex, while female (*XY*) is the heterogametic sex in birds.

The *Y* chromosome differs in morphology from the *X* chromosome, and generally it is smaller than the latter. *A region of the Y chromosome is homologous with a region of the X chromosome. A regular pairing between X and Y chromosomes occurs in this region; this ensures a regular separation of the X and Y chromosomes at AI so that they move to the opposite poles. In humans, the short arm of X chromosome* (X^s) *is homologous with the long arm of Y* (Y^L). *Therefore, the peculiar sex-linked inheritance is exhibited by the genes located in* X^L *only.*

Fertilization of an egg by a sperm having an *X* chromosome yields an *XX* zygote, which develops into a female. But zygotes produced by the union of an egg with a sperm having a *Y* chromosome produces an *XY* zygote; such zygotes produce males (Fig. 20.3).

20.5.2.4. XY Female, XX Male

This system of sex determination operates in birds, reptiles, some insects, *e.g.,* silkworm, etc. The scheme is essentially the opposite of that found in mammals etc. Here, the females have *XY* chromosome constitution. Thus females are the heterogametic sex as half of their eggs have an *X*, while the rest have a *Y* chromosome. The males of these species have two *X* chromosomes (*XX*), and hence are the homogametic sex. All the sperms produced by males, therefore, have one *X* chromosome.

Fertilization of an *X* containing egg with any sperm gives rise to an *XX* zygote, which develops into a male. An *XY* zygote is produced when a *Y* containing egg is fertilized by a sperm; such a zygote develops into a female (Fig. 20.3).

20.5.2.5. Diploid (2n) Female, Haploid (n) Male

This system of sex determination is found mainly in hymenoptera, *e.g.,* honeybees, ants, termites etc. In these species, the somatic chromosome number of females

is diploid, while that of males is only haploid. During spermatogenesis, all the *n* chromosomes of males regularly pass to a single pole at AI so that the opposite pole receives no chromosome at all. Thus all the sperms are regularly haploid. Normal meiosis during oogenesis produces all haploid eggs.

Fertilization of eggs produces diploid zygotes, which develop into diploid larvae. Ordinarily, such larvae give rise to workers, which are sterile females. But the diploid honeybee larvae fed on royal jelly develop into fertile females called ***queen***. Royal jelly contains honey, pollen and some substances (hormones?) secreted by workers. On the other hand, *unfertilized eggs develop parthenogenically to produce haploid larvae and, ultimately, fully fertile haploid males called* **drones** (Fig. 20.4).

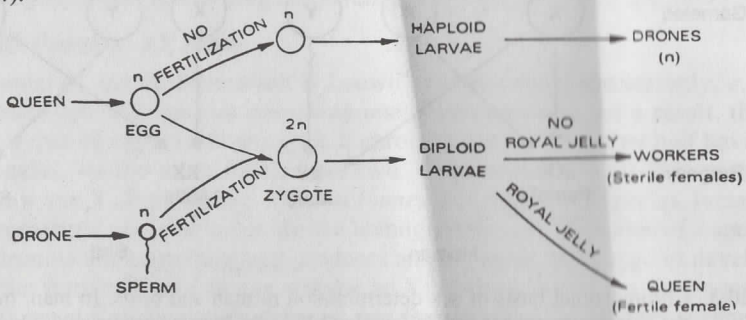


Fig. 20.4. Development of haploid males and diploid females in honey bee from unfertilized and fertilized eggs, respectively. Female larvae develop into fertile queen only when they are fed on royal jelly.

Genetic studies in a hymenopteran insect, *Bracon hebetor*, have clearly shown that *the males need not be haploid. Thus the maleness is not due to the haploid state per se.* It is suggested that sex determination in *Bracon* is governed by a single gene *X* with multiple alleles designated as X_a, X_b, X_c etc. Heterozygous state for this gene (*e.g.*, X_aX_b) produces femaleness, while maleness is the result of homozygous state for this gene. Ordinarily, males are haploid and hemizygous for this gene, which in effect is the same as homozygosity. But inbreeding in *Bracon* produces some diploid zygotes homozygous for an allele of the gene *X*; these zygotes develop into diploid males.

20.5.3. Single Genes Affecting Sex Differentiation

In many animals, single autosomal genes override the effect of sex chromosomes present in the individual. These genes are generally recessive, but in some cases they may be dominant. A classical example of such a gene is the autosomal recessive *transformer* (*tra*) gene of *Drosophila*. When *tra* is present in the homozygous state, it transforms *XX* zygotes to develop into males, which are sterile. The allele *tra* has no effect either in the males or when it is in the heterozygous state in females.

A similar recessive *testicular feminization* gene induces *XY* humans to develop breasts and vagina. Similar examples of single genes affecting sex differentiation are known in many other mammals. In addition, sex in many organisms is postulated to be governed by single genes, *e.g.*, gene *X* in *Bracon*. In many other cases, a single master gene, *e.g.*, *SRY* in man, *Sry* in mouse and *Sxl* in *Drosophila*, regulates sex differentiation. Gene *tra* of *Drosophila* is normal component of the series of genes

responsible for sex determination. Similarly, other such genes may also be components of the sex determination system.

20.6. MOLECULAR BASIS OF SEX DETERMINATION

It became clear quite early (by 1921) that determination of sex was rather complex at the genetic level. Two distinct models of sex determination have been discovered: (1) genic balance, and (2) Y-linked genes.

Genic Balance Theory

- Due to Bridges (1921) in case of *Drosophila.*
- Sex determined by sex index (= *X* / *A*):
  - (*i*) Index of 1.0 → Female.
  - (*ii*) Index of 0.5 → Male.
  - (*iii*) Index > 1.0 → Superfemale.
  - (*iv*) Index < 1.0 but > 0·5 → Intersex.
  - (*v*) Index < 0.5 → Supermale.
- *Gynandromophs* : part of body male, rest female; due to loss of one *X* from some cells of *XX* embryos.
- *X*/*A* ratio of 1.0 signals *Sxl* to produce active protein.
  - (*i*) Signal possibly from *sis-a* and *sis-b* genes.
  - (*ii*) *Sxl* protein causes *tra* to produce active protein.
  - (*iii*) *Tra* protein leads to production of *Tra*2 protein.
  - (*iv*) *Tra*2 protein causes *dsx* to produce a female-specific protein, which suppresses maleness; the embryos develops as female.
- *X*/*A* ratio of 0.5 causes *Sxl* to produce inactive protein:
  - (*i*) *tra* and *tra2* also produce inactive proteins.
  - (*ii*) As a result, *dsx* produces a male-specific protein that suppresses femaleness; embryo develops as male.

20.6.1. Sex Determination by Genic Balance

In *Drosophila*, man, mice etc., the presence of *XX* and *XY* chromosomes is ordinarily associated with femaleness and maleness, respectively. From the earlier discussion (section 20.5) it appears as if these specific chromosomes determine the sex of zygotes. But it should be emphasized that some specific genes located in these chromosomes are involved in sex determination. This conclusion was reached by Bridges who finally proposed his well recognized ***genic balance theory of sex determination*** in *Drosophila* in 1921.

Earlier, in 1916, Bridges discovered *XXY* females and *XO* males in *Drosophila* while studying the inheritance of *vermilion* eye gene located in the *X* chromosome. This clearly showed that *XX and XY chromosome constitutions are not essential for femaleness and maleness, respectively, and that Y chromosome did not play a role in sex determination.*

A little later, Bridges obtained triploid females. These females were mated with normal diploid males; as a result, they produced a number of aneuploid progeny (Table 20.1). The progeny from such a cross showed five different sex expressions. By correlating the sex of an individual with its chromosome constitution, Bridges developed the genic balance theory of sex determination. This theory fully explains

the sex determination mechanism in *Drosophila*, and most likely is applicable to birds as well.

TABLE 20.1

Sex Expression in *Drosophila* as a Function of the Ratio of the Number of X Chromosomes and the Number of Autosomal Sets Present in an Individual (Sex Index = X/A)

Ploidy	*Number of X chromosomes (= X)*	*Number of autosomal sets (= A)*	*Sex index (X/A)*	*Sex expression*
2*n*	3	2	3/2 = 1.5	Superfemale
3*n*	4	3	4/3 = 1.33	Superfemale
4*n*	4	4	4/4 = 1.0	***Female***
3*n*	3	3	3/3 = 1.0	***Female***
2*n*	2	2	2/2 = 1.0	***Female***
4*n*	3	4	3/4 = 0.75	Intersex
3*n*	2	3	2/3 = 0.67	Intersex
2*n*	1	2	1/2 = 0.5	***Male***
4*n*	2	4	2/4 = 0.5	***Male***
3*n*	1	3	1/3 = 0.33	Supermale

20.6.1.1. Sex Index

The ***genic balance theory*** states that the sex of an individual is determined by a balance between the genes for maleness and those for femaleness present in the individual. In *Drosophila, genes for maleness are presumed to be present in autosomes, while those for femaleness are considered to be located in the X chromosome. The genes for femaleness present in a single X chromosome are stronger than those for maleness present in one set of autosomes.* An autosomal set consists of the haploid set of autosomes. The strengths of genes for maleness and femaleness are so balanced that when the number of *X* chromosomes and that of autosomal sets is equal in an individual, it develops into a ***female***. An individual develops into a ***male*** only when the number of its *X* chromosomes is exactly half of the number of its autosomal sets. In essence, *the sex of an individual is determined by the ratio of the number of its X chromosomes and that of its autosomal sets;* this ratio is termed as ***sex index*** and is expressed as follows.

$$\text{Sex index} = \frac{\text{Number of } X \text{ chromosomes } (=X)}{\text{Number of autosomal sets } (=A)} = \frac{X}{A}$$

Individuals with the sex index of 1.0 are normal females, irrespective of whether they are *XX, XXX* or *XXXX* (*Table 20.1*). Similarly, ***flies having the sex index of 0.5 are normal males*** (*XO* diploid flies and *XXOO* tetraploid flies).

20.6.1.2. Intersex

Some flies have a sex index between 1.0 and 0.5; they develop into **intersexes,** *i.e.*, flies with a sex expression (both primary and secondary sex characters) intermediate between males and females. Intersexes are neither male nor female and are completely sterile. Some aneuploid flies have the sex index in excess of 1.0. These flies have more pronounced female characteristics than normal females, and are called ***superfemales***. Superfemales are generally weak and sterile, and often inviable. ***Supermale*** flies have a sex index of less than 0.5, are weak, sterile and often inviable; they show more pronounced secondary sex characteristics than

normal males. These explanations nicely fit the findings on various types of aneuploid flies obtained in various studies.

20.6.1.3. Gynandromorphs

In *Drosophila,* some individuals show male characteristics in a part of their body, while their remaining parts show the female phenotype; such individuals are known as ***gynandromorphs***. The male phenotype in gynandromorphs may be confined to a small part of the body, or may extend, to the maximum, to about one half of the body. In some flies, the male and female parts run longitudinally, while in some others they run transversely. *Gynandromorphs are always mosaics for X chromosome: the parts with male phenotype are always XO, while those with the female phenotype are XX.*

Gynandromorphs are believed to arise from XX zygotes. During embryonic development, in one or more cells one of the two X chromosomes does not pass to any pole at anaphase and, as a result, is lost. Consequently, one or more daughter cells having a single X chromosome are produced; these cells divide and give rise to the male parts of the gynandromorphs.

20.6.1.4. Mechanism of X/A Action

The *X/A* ratio (sex index) is measured very early during the development of an individual and a signal is converyed to an *X* - linked gene *Sxl* (*Sex lethal*). Gene *Sxl* is a master regulatory gene controlling sex differentiation. It derives its name from the fact that a loss of its function (due to a mutation) is lethal in females, while a gain of function is lethal in males. The main features of sex differentiation in *Drosophila* are briefly outlined below (Fig. 20·5).

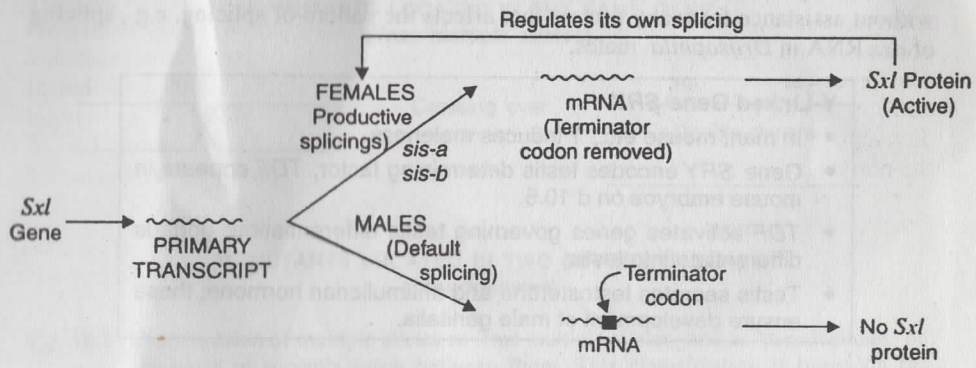


Fig. 20.5. Sex-specific differential splicing of *Sxl* primary transcript leads to the production of active *Sxl* protein in *Drosophila* females but not in males.

1. Gene *Sxl* is transcribed in both males and females.
2. When sex index of an embryo is 0·5, the RNA transcript of *Sxl* gene undergoes default splicing. The resulting mRNA does not code for a functional protein. *In the absence of functional Sxl protein, the embryo develops into a male.*
3. When the sex index of an embryo is 1·0, a signal induces alternative splicing of the RNA transcript of gene *Sxl*. The resulting mRNA codes for

a functional Sxl protein. Sex - linked genes *sis-a* and *sis-b* seem to be involved in the signal that leads to productive splicing of *Sxl* RNA.

4. Protein Sxl causes productive splicing of RNA transcribed from gen *tra* . Tra protein has an active feminizing effect as described below. (In male embryos, RNA transcript of *tra* gene undergoes default splicing as there is no Sxl protein. The resulting mRNA does not produce any protein.)
5. Tra protein, in turn, causes productive splicing of the RNA transcribed from *tra*2 gene; the resulting mRNA yields Tra2 protein. (In males *tra*2 RNA undergoes nonproductive default splicing so that male embryos do not have Tra2 protein.)
6. Tra2 protein causes females - specific splicing of the RNA from gene *dsx*. The protein encoded by this mRNA is a female- specific Dsx protein, which causes the supression of male genes. Dsx protein probably acts together with Ix protein (encoded by gene *ix*, *intersex*). Therefore, *XX* embryos develop as females.
7. In case of males, Tra2 protein is absent. Therefore, the RNA product of *dsx* gene undergoes a male- specific splicing; the mRNA so obtained gives rise to a male-specific Dsx protein that blocks female sexual development. Therefore, *XY* and *XO* embryos develop as females.

Splicing is the process in which noncoding sequences (***introns***) are removed from RNA transcripts of eukaryotic genes and all the coding sequences (***exons***) present in the transcript are pieced together in the correct order to yield mRNA molecules. A ***productive splicing*** yields mRNA molecules that are translated into functional proteins. In contrast, ***nonproductive splicing*** generates mRNA molecules that do not produce functional proteins. In case of ***default splicing***, splicing occurs without assistance from any protein that affects the pattern of splicing, e.g., splicing of *dsx* RNA in *Drosophila* males.

Y-Linked Gene *SRY*

- In man, mouse etc., *Y* induces maleness.
- Gene *SRY* encodes testis determining factor; *TDF* appears in mouse embryos on d 10.5.
- *TDF* activates genes governing testis differentiation; gonads differentiate into testis.
- Testis secretes testosterone and antimullerian hormone; these ensure development of male genitalia.

20.6.2. Sex Determination by Y-Linked Genes

In mammals, axolotl (an Amphibian) and some plants, *e.g., Melandrium,* the *Y* chromosome is essential for the development of maleness. In humans *XO, XX, XXX* and *XXXX* individuals develop the female phenotype. But in the presence of a single *Y* chromosome, these individuals, *i.e.., XY, XXY, XXXY* and *XXXXY*, develop the male phenotype. *Thus a single Y chromosome is sufficient to overcome the effects produced by four X chromosomes. But normal human females and males are produced only by XX and XY chromosome constitutions, respectively.* The *XO* condition produces Turner's syndrome, while *XXY* leads to the development of Klinefelter's syndrome.

20.6.2.1. H-Y Antigen

The male determining capacity in humans is known to be located in the short arm of *Y*, (Y^s). A deletion for Y^s permits the development of normal female phenotype even in the presence of such a *Y* chromosome. In most mammals, the expression of an antigen, H-Y antigen, is closely correlated with testis development. The gene (or genes) governing the presence of H-Y antigen is located in the short arm of the *Y* chromosome. However, the H-Y antigen is not directly involved in sex differentiation.

20.6.2.2. Testis Determining Factor (TDF)

A *Y*- linked gene *SRY* encodes TDF, which directs differentiation of testis. In mouse *XY* embryos, TDF is produced around day 10.5, while testis differentiation begins on day 11.5 of embryo development. *SRY* is named after *s*ex *d*etermining *r*egion of *Y*; this notation is used in man, while in mouse *Sry* symbol is used. Sequences homologous to *SRY* have been found in many species like cattle, rabbit, horse, tiger, marsupials etc.

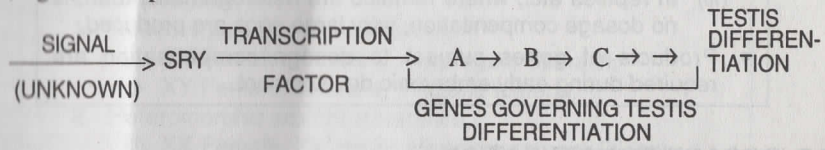


Fig. 20.6. Schematic representation of activation of genes governing testis differentiation by S_{RY}. The signal that activates S_{RY} is unknown.

SRY encodes a transcription factor, which binds to DNA and activates the transcription of genes involved in testis differentiation (Fig. 20·6). The identity of genes involved in testis development is largely unknown; some such genes are *SRVX* (sex reversal *X*; *X* - linked), *SRA*1 (sex -reversal autosomal; chromosome 17 short arm) and a gene on chromosome 9 (long arm). (Mutations in these genes cause sex reversal of *XY* individuals.). Testis secretes testosterone and antimullerian hormone, which together ensure the development of male body type. Testosterone acts via the androgen receptor encoded by an *X*-linked gene. A mutation in this gene leads to the development of female genitalia in *XY* individuals having normal testes, *i.e.*, testicular feminization (Section 20.5.3). Our knowledge of sex differentiation in human beings is summarised in Fig. 20.7.

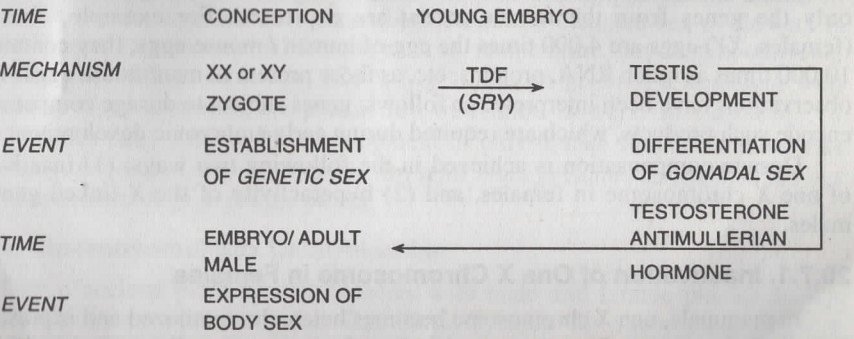


Fig. 20·7. A summary of sex differentiation in human beings; differentiation of only males shown.

20.6.2.3. Hermaphrodites

An individual having ovarian follicles and testicular tubules in the same or separate gonads is called ***hermaphrodite***. Some *XX* and *XY* individuals are true hermaphrodites. The *XY* hermaphrodites are usually mosaics for gonad differentiation. This mosaic condition, at least in some cases, appears due to a mosaic expression of the gene *SRY* so that TDF is produced in some tissues only.

Dosage Compensation

- In mammals etc., females have 2 *X* chromosomes, and males have 1.
- But expression of *X*-linked genes is comparable in males and females; it is achieved as follows.
  - (*i*) One *X* chromosome becomes heterochromatinized (Barr body) in females (d 12 of embryo development) in mammals.
  - (*ii*) In *Drosophila*, *X*-linked genes are hyperactive in males.
  - (*iii*) In reptiles etc., where females are heterogametic, there is no dosage compensation; very large eggs are produced.
- Products of genes subject to dosage compensation are required during early embryonic development.

20.7. DOSAGE COMPENSATION

In man, mouse, *Drosophila* etc., females (*XX*) have two *X*-chromosomes, while their males (*XY*) have only *X* chromosome. Therefore, females have two copies of the *X*-linked genes, while males have only one copy of each such gene. But the intensities of the phenotypes, the amounts of the concerned enzymes, and even the amount of mRNAs produced by the *X*-linked genes in the males are equal to those in females. This phenomenon, is called ***dosage compensation***; it regulates the expression of *X*-linked genes in such a manner that they produce comparable effects in males and females even though they have different number of copies (= dosage) of these genes.

Dosage compensation occurs in such species where male is the heterogametic sex. In such species, where females are the heterogenetic sex, *e.g.*, reptiles, birds etc. there is no dosage compensation. In these species, very large eggs are produced, and the timing of activation of genes in the embryos seems to be so adjusted that initially only the genes from the female parent are expressed. For example, *Xenopus* (females, *XY*) eggs are 4,000 times the egg of human / mouse eggs; they contain Ca. 10,000 times as much RNA, proteins, etc. as those present in man/mouse eggs. These observations have been interpreted as follows: genes subject to dosage compensation encode such products, which are required during early embryonic development.

Dosage compensation is achieved in the following two ways: (1) inactivation of one *X* chromosome in females, and (2) hyperactivity of the *X*-linked genes in males.

20.7.1. Inactivation of One X Chromosome in Females

In mammals, one *X* chromosome becomes heterochromatinized and is present as Barr body in somatic interphase nuclei. Barr body was first discovered in 1949 by Barr and Bertram. It first appears on day 12 of embryo development. Barr body

supports only 18% RNA synthesis of normal, replicates the last during *S* phase, and is not found in premeiotic cells and gametes. The number of Barr bodies in an individual is equal to *X*- (*A*/2), where *X* is the number of *X* chromosomes and A is the number of autosomal sets. The inactivation of *X* chromosomes may be random or non-random.

20.7.2. Hyperactivity of X-Linked Genes in Males

In *Drosophila*, the genes present in the single *X* chromosome of males produce as much mRNA as those present in the two *X* chromosomes of females. Clearly, the *X*-linked genes are more active (hyperactive) in males; this was conclusively shown by Mukherjee and later confirmed by other workers. This hyperactivity depends on *X/A* ratio of flies and is specific for, *X*-linked genes only.

Sex Determination in Plants

- *Environmental* : in papaya, melons, cucumber etc.
- *Chromosomal* :
  1. Homomorphic sex chromosomes:
     - (*i*) XX Female, XY male; *Asparagus, Spinach.*
     - (*ii*) XY Female, XX male; *Fragaria elateria.*
  2. Heteromorphic sex chromosomes:
     - (*i*) XX Female, XY male; *Melandrium* (now *Silene*).
     - (*ii*) XX Female, XY_1Y_2 male; *Rumex acteosa,* etc.
     - (*iv*) $X_1X_1X_2X_2$ female, $X_1X_2Y_1Y_2$ male; *Humulus lupulus.*
- *Geneic : Ecballium, Mercurialis,* etc.

20.8. SEX DETERMINATION IN PLANTS

The mechanisms of sex determination in plants are essentially similar to those in animals. These mechanisms are also of three types: (1) environmental, (2) chromosomal and (3) monogenic (Table 20.2).

20.8.1. Environmental Sex Determination

In some plants, either sex determination is due to the environment or it is greatly affected by the environment. *Equisetum* plants grown under optimum conditions develop as females, while those grown under adverse conditions become males. In many other plants, *e.g.*, melons, cucumber, *Cannabis* etc., the sex of flowers is affected by many environmental factors, such as, temperature, day-length, ethylene, gibberellic acid (GA_3), some ions (Ca^{2+}, Mg^{2+}), etc. In general, a treatment with ethylene or GA_3 promotes the production of female flowers. In *Cannabis,* a treatment with GA_3 induces the development of only female flowers in otherwise male plants. *In general, the environment has a far more pronounced effect on the sex expression in male than in female plants.*

20.8.2. Chromosomal Sex Determination

Many dioecious plant species (species with male and female plants) do not show distinct sex chromosomes. But some species do show clear-cut sex chromosomes. The chromosomal mechanisms of sex determination may be grouped into the following categories: (1) homomorphic sex chromosomes, (1*a*) *XX* female, *XY* male, (1*b*) *XY* female, *XX* male, (2) heteromorphic chromosomes, (2*a*) *XY*

females, XY male, (2*b*) XX female, XY_1Y_2 male and (2*c*) $X_1X_1X_2X_2$ female, $X_1X_2Y_1Y_2$ male (Table 20·2).

TABLE 20·2
Mechanisms of Sex Determination in Plants

Mechanism	*Example species*
Environmental	*Equisetum* sp.
Chromosomal	
1. Homomorphic chromosomes	
1*a*. Heterogametic male (XX, XY)	*Asparagus, Spinach*
1*b*. Heterogametic female (WZ, ZZ or XY, XX)	*Fragaria elateria*
2. Heteromorphic chromosomes (Heterogametic male)	
2*a*. Active Y (XX; XY)	*Silene latifolia* (Syn., *Melandrium*)
2*b*. X/autosome balance (XX, XY_1Y_2)	*Rumex acetosa*
2*c*. $X_1X_1X_2X_2$, $X_1X_2Y_1Y_2$	*Humulus lupulus*
Genic	
1. Single locus	*Ecballium elaterium*
2. Multiple loci	*Mercurialis annua*

20.8.2.1. Heteromorphic Sex Chromosomes.

In such plant species, the X and Y chromosomes show both morphological and functional differentiation so that they are identifiable cytologically. As listed above, three distict sex determination mechanisms are known under this category; in all the cases, males are the heterogametic sex.

Sex Determination in *Melandrium* (*Silene*)

- Y chromosome determines maleness; X specifies femaleness.
- A single Y fully overcomes the effect of 3X chromosomes.
- Y chromosome has 4 functional regions:
  - (*i*) *Segment I:* Terminal; suppresses femaleness.
  - (*ii*) *Segment II*: Located next to segment *I*; promotes maleness.
  - (*iii*) Segment *III*: Next to region *II*; governs male fertility.
  - (*iv*) Segment *IV*: Terminal; homologous to segment of X chromosome.
  - (*v*) The remaining portion of X chromosome determines femaleness.

20.8.2.1.1. XX Female, XY Male

This system is the most common in plants, and is found in *Cannabis, Melandrium* (now *Silence*) *Rumex acetosella, Salix, Bryonia* etc. In these species, female plants have two X chromosomes, hence all the egg cells have one X chromosome. Male plants have one X and one Y chromosome. Therefore, half of the pollen grains have one X chromosome, while the remaining pollen grains have a Y chromosome. Random union between male and female gametes produces two types of zygotes: 50% of the zygotes are XX and give rise to female plants, and the remaining 50% zygotes are XY and produce male plants. This system operates in the same manner as that found in mammals (Fig. 20.3).

In *Melandrium*, the Y chromosome determines maleness, while X chromosome specifies femaleness (Fig. 20.8). A single Y chromosome is able to produce male flowers even in the presence of four X chromosomes ($XXXXY$), but such plants do produce some bisexual flowers. However, $XXXY$ plants are fully male and fertile.

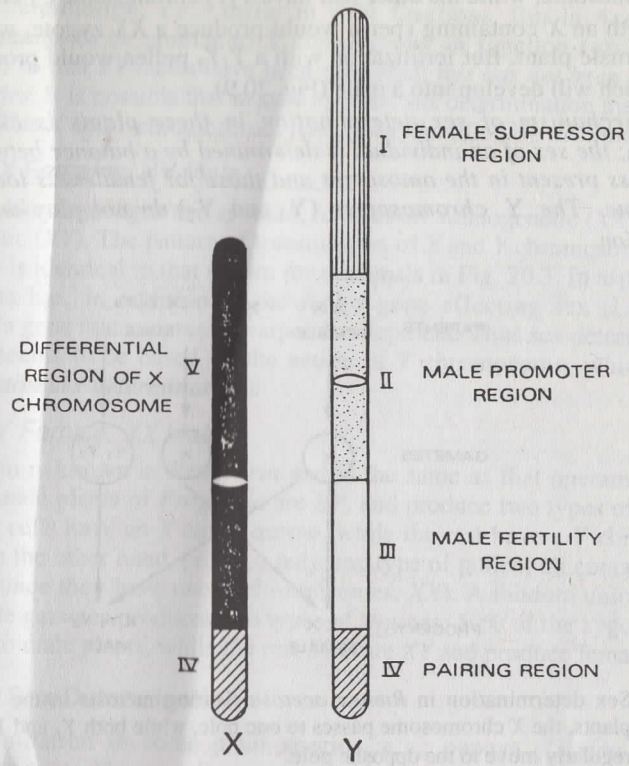


Fig. 20.8. Organisation of X and Y chromosomes of *Melandrium* (now, *Silene*) in respect of their roles in sex-differentiation. A loss of segment I from the Y chromosome produces bisexual flowers, that of region II results in a female plant, while that of segment III results in a male sterile plant with abortive anthers.

The Y chromosome of *Melandrium* is the largest chromosome of its genome, It has four functionally distinct segments. The ***first segment*** contains female suppressor genes and is located at one end of the Y chromosome. A deletion for this segment produces hermaphrodite flowers on XY plants. The ***second region,*** lying next to the first, has genes that promote maleness, *i.e.,* initiate the development of anthers. The ***third segment*** is located next to the second, and contains genes governing male fertility and anther maturation. A deletion for this region produces male sterile plants with abortive anthers. The ***fourth segment,*** located at the other end of the Y chromosome, is homologous to a segment of the X chromosome (Fig. 20.8). Pairing occurs between X and Y chromosomes in this homologous region; this ensures a regular separation of X and Y chromosomes at AI of meiosis. A large part (segment V) of the X chromosome is nonhomologous to the Y chromosome; this region has genes governing the development of female reproductive organs.

20.8.2.1.2. XX Female, XY_1Y_2 Male

This system is found in *Rumex acetosa* and *Humulus japonicus*. Female plants have two X chromosomes, hence all their egg cells have one X chromosome. But male plants are XY_1Y_2, and produce two types of pollen: one-half of the pollen have

one *X* chromosome, while the other half have Y_1Y_2 chromosomes. Fertilization of an egg cell with an *X* containing sperm would produce a *XX* zygote, which will give rise to a female plant. But fertilization with a Y_1Y_2 pollen would produce an XY_1Y_2 zygote, which will develop into a male (Fig. 20.9).

The mechanism of sex determination in these plants is akin to that in Drosophila*; the sex of an individual is determined by a balance between the genes for maleness present in the autosomes and those for femaleness located in the* X *chromosome. The* Y *chromosomes* (Y_1 and Y_2) *do not play any role in sex determination.*

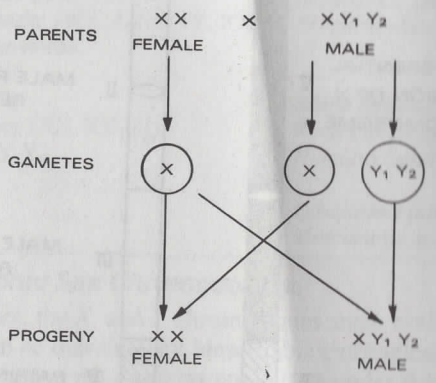


Fig. 20.9. Sex determination in *Rumex acetosa.* During meiosis in the PMC's of male plants, the *X* chromosome passes to one pole, while both Y_1 and Y_2 chromosomes regularly move to the opposite pole.

20.8·2·1·3. $X_1X_2X_2X_2$ Female, $X_1X_2Y_1Y_2$ Male

This system is found in some strains of *Humulus lupulus*.. Females are homogametic and all eggs produced by them are X_1X_2. In contrast, males are heterogametic and produce two types of sperms, *viz.*, X_1X_2 and Y_1Y_2. Fertilization of an egg with an X_1X_2 sperms will yield a females ($X_1X_1X_2X_2$), while that with a Y_1Y_2 sperm will produce a male ($X_1X_2Y_1Y_2$) progeny (Fig. 20.10).

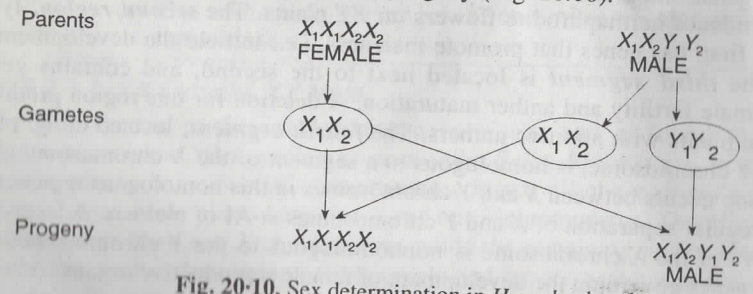


Fig. 20·10. Sex determination in *Humulus lupulus.*

20.8.2.2. Homomorphic Sex Chromosomes

In several plant species, *X* and *Y* chromosomes are identical in morphology, *e.g.*, in *Asparagus* and *Spinach*. The *X* and *Y* chromosomes, however, do show some

functional differentiation, *i.e.*, differ in their functions. But in *Asparagus*, *Y* chromosome seems to be similar to *X* chromosome in function (except for sex differentiation) so that *YY* males have been produced; *this has not been achieved in any other species*. It is possible that in such species, sex determination may be due to a single gene. The *X* and *Y* chromosomes have been identified by trisomic analysis.

20.8.2.2.1. XX Females, XY Male

In species like *Asparagus* and *spinach*, females are homogenetic (*XX*) and males are heterogenetic (*XY*). The pattern of transmission of *X* and *Y* chromosomes and the sex of progeny is identical to that shown for mammals in Fig. 20.3. In asparagus, the Y chromosome has, in addition to the major gene affecting sex (*i.e.*, stamen development), a gene that suppresses carpel development. Thus sex determination in these species seems to be based on the action of *Y* chromosome, which induces maleness (***Y-active sex determination***).

20.8.2.2.2. XY Female, XX Male

This system is known in *Fragraria* and is the same as that operating in birds (Fig. 20.3). Female plants of *Fragraria* are *XY*, and produce two types of egg cells: half of the egg cells have an *X* chromosome, while the rest have a *Y* chromosome. Male plants, on the other hand, produce only one type of pollen, all containing an *X* chromosome (since they have two *X* chromosomes, *XX*). A random union between male and female gametes produces two types of zygotes: 50% of the zygotes are *XX* and develop into male plants, while the rest 50% are *XY* and produce female plants.

20.8.3. Genic Sex Determination

Sex determination in some plant species, *e.g.*, papaya (*Carica papaya*), *Ecballium elaterium, Mercurialis annua,* etc. is postulated to be governed by a single gene. *In papaya, a single gene with three alleles* (m, M_1 *and* M_2) *is suggested to control the sex of an individual. Female plants are homozygous* mm, *while male plants are heterozygous* M_1m; *the heterozygote* M_2m *produces hermaphrodite condition. Genotypes* M_1M_1, M_1M_2 *and* M_2M_2 *are inviable* (Table 20.3).

Matings between females (*mm*) and males (M_1m) produce 50% male (M_1m) and 50% female (*mm*) progeny. Similarly, a cross between a female (*mm*) and a hermaphrodite (M_2m) plant would yield 50% hermaphrodite (M_2m) and 50% female plants (*mm*). Selfing of a hermaphrodite plant produces 2/3 hermaphrodite (M_2m) and 1/3 female (*mm*) plants; about 1/4 of the zygotes are M_2M_2 and do not survive.

A single gene control of sex in found in spitting cucumber (*Ecballium elaterium*). In this species, gene *a* has 3 alleles, *viz.*, a^d for femaleness, a^D specifying maleness, and a^+ leading to monoecy (male and female flowers are produced on the same plant). The dominance relationship among these alleles in as follows : $a^+ > a^D > a^d$. Any given variety of *E. elaterium* is either monoecious or dioecious. In a dioecious variety, female plants have the genotype a^da^d, while male plants are $a^D\,a^d$. Therefore, progeny from male × female matings are 1 male (a^Da^d) : 1 female (a^da^d) . Plants of a monoecious variety are $a^+\,a^+$ in genotype. Therefore, selfing produces all dioecious ($a^+\,a^+$) plants. But when monoecious plants are crossed with, say, female plants, 2% of female plans are also produced in the progeny (the remaining plants are monoecious, $a^+\,a^d$). This indicates that genes other than *a* also affect sex determination in this species (Table 20·4).

TABLE 20.3
Postulated Monogenic sex Determination in Papaya; the Gene *m* has Three Alleles, *viz.*, *m*, M_1 and M_2

Genotype	*Survival*	*Sex expression*
mm	Vital	Female
M_1m	Vital	Male
M_2m	Vital	Hermaphrodite
M_1M_1, M_1M_2 and M_2M_2	Lethal (all die)	
	Various Crosses	
Female × male ($mm \times M_1m$)		1/2 Male (M_1m), 1/2 female (mm)
Female × hermaphrodite ($mm \times M_2m$)		1/2 Hermaphrodite ($M_2\,m$), 1/2 female ($m\,m$)
Hermaphrodite ($M_2\,m$) selfed		1/4 $M_2\,M_2$ (inviable), 1/2 hermaphrodite ($M_2\,m$), 1/4 female (mm)

TABLE 20.4
Single Gene Sex Determination in Spitting Cucumber (*E. elaterium*)

Genotype	*Phenotype*
$a^d\,a^d$	Female
$a^D\,a^d$	Male
a^+a^+ (also a^+a^d, a^+a^D)	Monoecious
Crosses	
Female (a^da^d) × Male (a^Da^d)	1 Female (a^da^d) : 1 Male (a^Da^d)
Female (a^da^d) × Monoecious (a^+a^+)	Monoecious (a^+a^d) (~ 2% female plants)*

*Genes other than *a* also affect sex expression.

In some plant species, more than one gene is involved in sex determination. For example, in annual mercury plant (*Mercurialis annua*), sex is governed by three unlinked genes *A*, B_1 and B_2. The principle of sex determination in *M. annua* is as follows.

1. ***Male plants*** always have at least one dominant allele of gene *A*. In addition, they also have at least one dominant allele of either gene B_1 or B_2 or of both. Therefore, male plants may have the following genotypes: $AA\ B_1b_1\ B_2b_2$, $Aa\ B_1b_1\ B_2b_2$, $AA\ B_1B_1\ b_2b_2$, $Aa\ B_1B_1\ b_2b_2$, $Aa\ B_1b_1\ b_2b_2$, $AA\ b_1b_1\ B_2b_2$, $Aa\ b_1b_1B_2B_2$ and $Aa\ b_1b_1B_2b_2$.
2. ***Female plants*** are homozygous recessive at any two of the three loci. Therefore, female plants may be of the following genotypes: $AA\ b_1b_1b_2b_2$, $aa\ B_1B_1\ b_2b_2$ and $aa\ b_1b_1\ B_2B_2$.

QUESTIONS

1. List the various mechanisms of sex determination in animals. Describe the environmental and monogenic sex determination.

2. List the various mechanisms of sex determination in animals. Describe the mode of sex determination in *Drosophila* assuming chromosomal as well as genic balance as the basis of sex determination.
3. 'Sex determination is not the direct result of *X* and *Y* chromosomes themselves. But it is due to specific sex genes present in them'. Discuss this statement with the help of suitable examples (*e.g.*, in *Drosophila* and in mammals).
4. Describe the different modes of sex determination in plants.
5. Write short notes on the following:
   (*i*) *XX* female, *XO* male,
   (*ii*) sex chromosomes,
   (*ii*) autosomes,
   (*iv*) *Y* chromosome,
   (*v*) intersexes,
   (*vi*) gynandromorphs.

21

Sex Linkage

21.1. INTRODUCTION

Ordinarily, F_1 and F_2 generations from reciprocal crosses yield identical results, and it does not matter if the female or the male parent had the recessive character. Further, both males and females in the progeny show identical ratios. But a notable exception to this general rule is produced due to the phenomenon of ***sex linkage***. *This is an association during inheritance of a character with sex of the progeny. Sex linkage is the consequence of a gene being located in the* X *or sex chromosome.* The *X* chromosome occurs in different numbers in the two sexes (*XX* in one and only one *X* in the other).

Sex Linkage
- Association during inheritance of a character with the sex of progeny.
- Due to location of the gene on *X* chromosome.
- First explained by Morgan (1910) for white eye in *Drosophila.*

The earliest known case of sex linkage is that of haemophilia in human beings. Persons afflicted by hemophilia bleed profusely even from smallest cuts since their blood does not clot on exposure to air. Jews had recognised even before 600 A.D. that haemophilia was a hereditary defect and that it was transmitted through females, although ordinarily only males were haemophilic. Other similar examples have also been described. But a clear-cut explanation for this phenomenon was presented by Morgan in 1910 for the *white* eye (*w*) gene of *Drosophila.*

21.2. INHERITANCE OF *WHITE* EYE IN *DROSOPHILA*

In a pedigree culture of normal dull red-eyed *Drosophila,* Morgan observed a single white-eyed male. When this white-eyed male was mated with red-eyed females, F_1 flies had red eyes. In F_2, on an average, 3 flies had red eyes and 1 had white eyes (Fig. 21.1). These results reveal that white eye is due to a single recessive gene. *But when the* F_2 *individuals were classified on the basis of their sex as well as eye colour, a peculiar picture emerged. All the* F_2 *females had red eyes, but half of the males had red and the other half had white eyes. It appears that the eye colour of a fly in* F_2 *depended on its sex.*

But when white-eyed female flies were mated with red-eyed males, *half the flies in* F_1 *were red-eyed and the remaining half had white eyes. In* F_2 *from this reciprocal cross, red-eyed and white-eyed flies were obtained in a 1 : 1 ratio, which is the same as in the* F_1 *generation.* These findings are very puzzling since *segregation for eye colour was obtained in* F_1 *and a 1 : 1 ratio was found in* F_2.

Further, the classification of F_1 and F_2 flies on the basis of their sex as well revealed that *all the* F_1 *males were white-eyed, while the females were red eyed.* However, in F_2 both male and female flies showed the ratio 1 red : 1 white (Fig. 21.1).

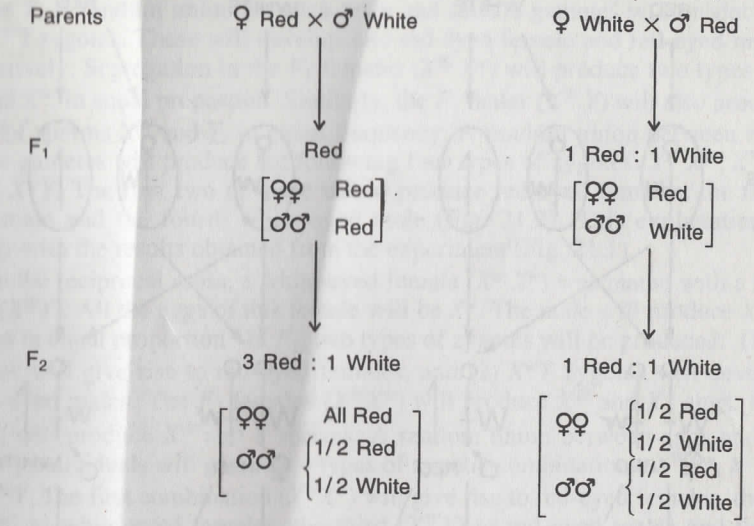


Fig. 21.1. The peculiar pattern of inheritance of white eye in *Drosophila.* Note that only F_2 males have white eyes in the first cross, while a 1 : 1 ratio is obtained in F_1 as well as F_2 in the reciprocal cross.

Morgan correctly reasoned that the *white* eye gene was located in the *X* chromosome of *Drosophila.* Therefore, the inheritance pattern of this gene closely parallels that of the *X* chromosome (Fig. 21.2). It was also assumed that the *Y* chromosome does not carry an allele for this gene. Accordingly, female files will have two copies of *w* gene (*WW, Ww* or *ww*). But males will have only one copy (*W* or *w*). Thus male *Drosophila* will be ***hemizygous*** (having only one copy of a gene) for *w* gene. Therefore, a single copy of the recessive allele *w* will express itself in the males.

White* Eye of *Drosophila
- Unusual inheritance pattern:
  1. (Red ♀ × white ♂ → F_1 ♂♂ and ♀♀ Red → F_2: ♀♀ Red, 1/2 ♂♂ Red, 1/2 ♂♂ white (Ratio: 3 Red : 1 White).
  2. (White ♀ × Red ♂ → F_1: ♀ ♀ Red, ♂♂ white (Ratio: 1 Red : 1 white) → F_2: 1/2 ♀♀ Red, 1/2 ♂♂ whit (Ratio : 1 Red: 1 White).
- Eye colour inheritance pattern the same as that of *X* chromosome
- *Conclusion* : *White* eye gene located in *X* chromosome.
- *First clear-cut demonstration of a specific gene being located on a particular chromosome.*

The genotypes of red-eyed females and white-eyed males may be written as $X^W X^W$ and $X^w Y$, respectively. The location of *w* gene in the *X* chromosome is

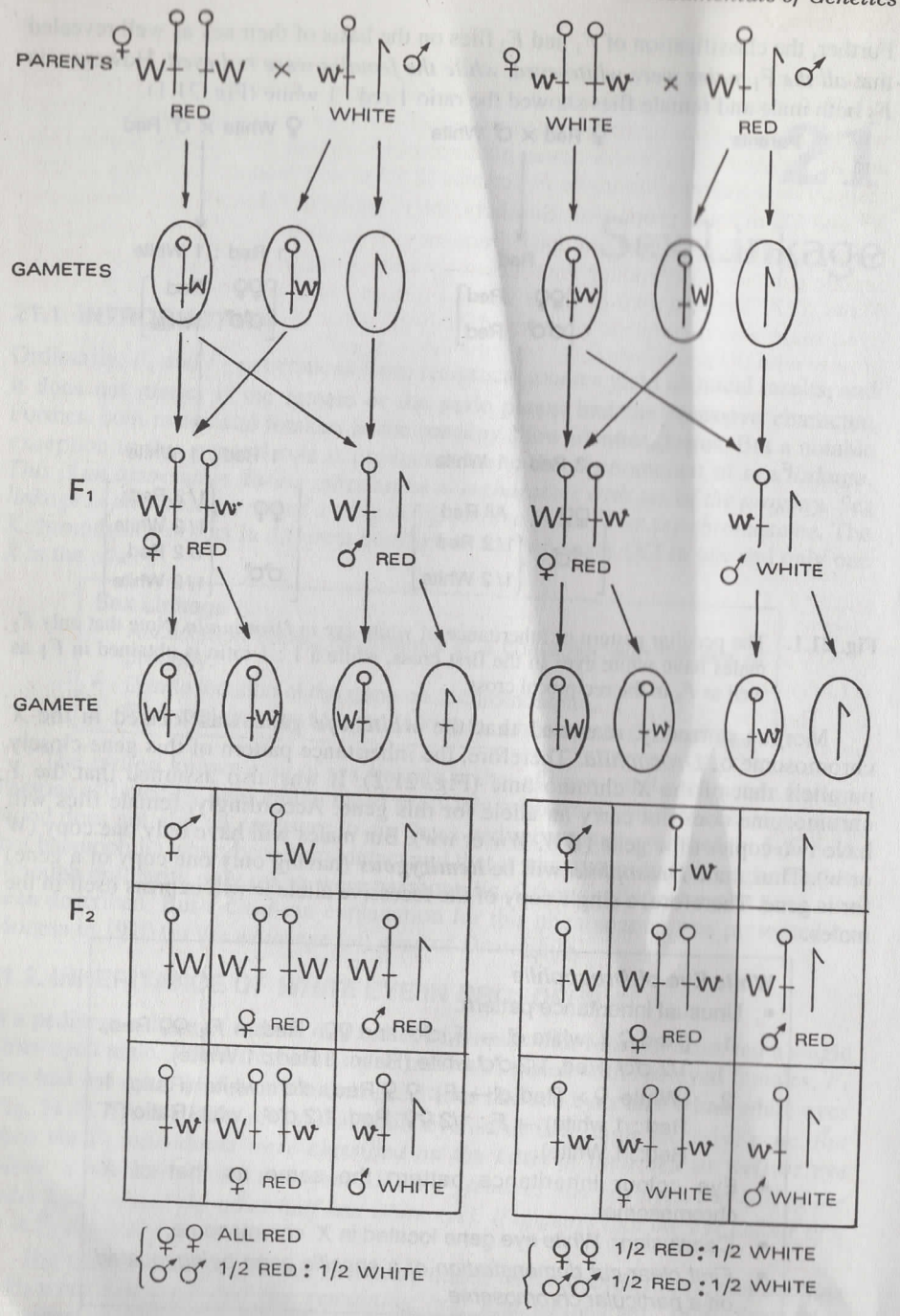


Fig. 21.2. Explanation of the inheritance of white eye in *Drosophila* by assuming that the *white* eye gene is located in the *X* chromosome

indicated by writing it as a superscript of *X*. All the eggs of the red-eyed female will be X^W. But half of the sperms of the white-eyed male will be X^w and the other half will be *Y*. A random union between male and female gametes will produce $X^W X^W$ and $X^W Y$ zygotes. These will develop into red-eyed female and red-eyed male flies, respectively. Segregation in the F_1 females ($X^W X^w$) will produce two types of eggs, X^W and X^w, in equal proportion. Similarly, the F_1 males ($X^W Y$) will also produce two types of sperms X^W and *Y*, in equal frequency. A random union between male and female gametes will produce the following four types of zygotes: $X^W X^W$, $X^W X^w$, $X^W Y$ and $X^w Y$. The first two of these would produce red-eyed females, the third red-eyed male and the fourth white-eyed male (Fig. 21.2). This explanation agrees closely with the results obtained from the experiment (Fig. 21.1).

In the reciprocal cross, a white-eyed female ($X^w X^w$) was mated with a red eyed male ($X^W Y$). All the eggs of this female will be X^w. The male will produce X^W and Y sperms in equal proportion. In F_1, two types of zygotes will be produced: (1) $X^W X^w$ zygotes will give rise to red-eyed females, and (2) $X^w Y$ zygotes will develop into white-eyed males. The F_1 females ($X^W X^w$) will produce X^W and X^w eggs. F_1 males ($X^W Y$) will produce X^w and *Y* sperms. A random union between eggs and sperms from F_1 individuals will generate 4 types of zygotic combinations: $X^W X^w$, $X^w X^w$, $X^W Y$ and $X^w Y$. The first combination ($X^W X^w$) will give rise to red-eyed females, the second ($X^w X^w$) to white-eyed females, the third ($X^W Y$) to red-eyed males and the fourth ($X^w Y$) to white-eyed males (Fig. 21.2). Clearly, these are the same as the results actually obtained (Fig. 21.1).

Thus *the hypothesis that* w *gene is located in the* X *chromosome and that it has no allele in the* Y *chromosome is adequate to explain the peculiar inheritance pattern of the white eye trait in* Drosophila. *This was the first conclusive demonstration that a specific gene was located in a specific chromosome of an organism.*

Nondisjunction of *X* Chromosome
- The two homologues of *X* do not separate at AI; they pass to the same pole producing *XX* gamete.
- Discovered by Bridges (1916) in *Drosophila* for *vermilion* eye colour (sex-linked).
- Produces unusual inheritance pattern, *viz.*, vermilion ♀ *X* red ♂ → F_1: vermilion ♀♀ and Red ♂♂. (Normally, F_1 should be, ♀♀ red, ♂♂ vermilion).
- Confirms location of *vermilion* gene on *X* chromosome.

21.3. NONDISJUNCTION OF *X* CHROMOSOME

In 1916, while studying the inheritance of *vermilion* eye colour in *Drosophila*, Bridges obtained some peculiar data. Vermilion eye colour is produced by a sex-linked recessive gene *v*. This gene shows the same inheritance pattern as *w* (Figs. 21.1, 21.2). Among the F_1 progeny from a cross between vermilion females and red males, some females had vermilion eyes, while some males had red eyes (Fig. 21.3). *The appearance of vermilion females and red males in the* F_1 *generation of this cross was contrary to the expectations based on sex linkage.* However, a vast majority of the progeny females had, as expected, red eyes and the males had vermilion eyes.

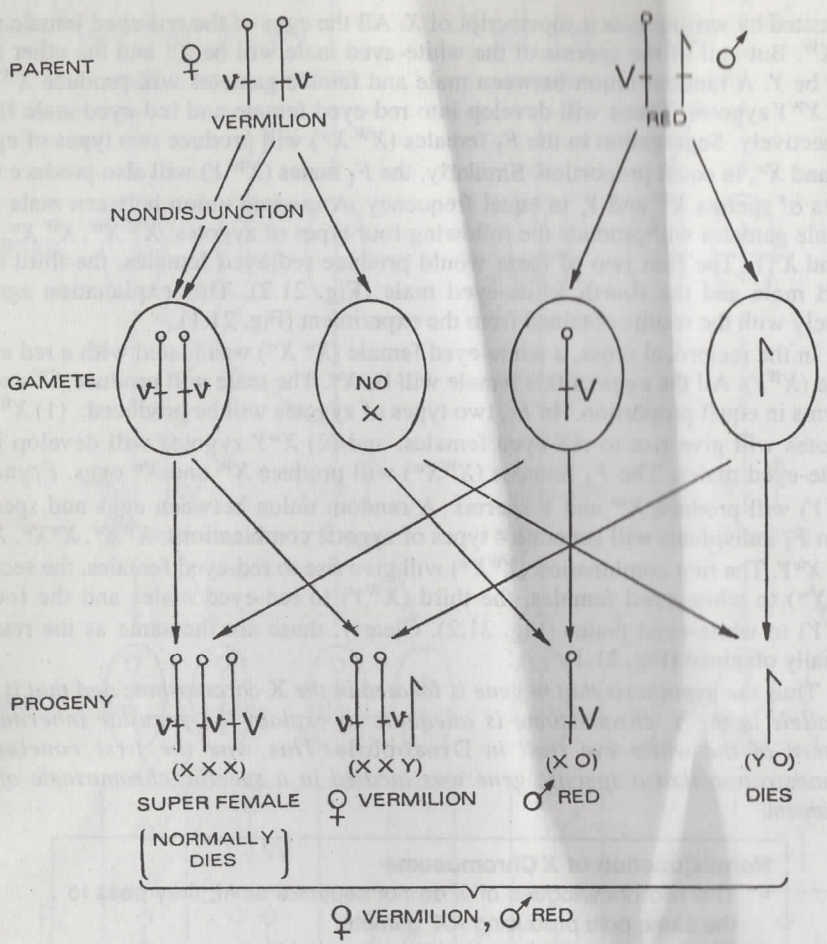


Fig. 21.3. Occasional unusual transmission of *vermilion* eye gene in *Drosophila* due to the non-disjunction of *X* chromosome. Note the transmission of *v* allele from male to male

To explain these data, Bridge postulated that during oogenesis, in some of the oocytes of the vermilion-eyed females (X^vX^v), the *X* chromosomes failed to separate at AI and they both moved together to the same pole. As a result, the opposite pole in such oocytes did not recessive any *X* chromosome. Such an irregular distribution of the homologous chromosomes at AI is known as ***nondisjunction***. Since in this case irregular distribution of *X* chromosome is postulated, it is referred to as ***nondisjunction of X chromosome.*** From those oocytes of the vermilion females (X^vX^v), which show nondisjunction of the *X* chromosome, two types of unusual eggs will be obtained: (1) half of the eggs will have two *X* chromosome and will be X^vX^v, and (2) the other half will be *O* as they would receive no *X* chromosome. Red-eyed males (X^VY) will produce X^V and *Y* sperms.

A random union between these male and female gametes will produce the following four types of zygotic combinations: $X^vX^vX^v$, X^vX^vY, X^vO and OY. The first

and the fourth combinations ($X^vX^vX^v$ and OY) will normally die. The second combination (X^vX^vY) will produce vermilion female, while the third (X^vO) will yield red-eyed male. Since nondisjunction of *X* chromosome may be expected to occur only occasionally, such unusual males and females will be recovered in low frequencies. Bridges *concluded that the unusual vermilion-eyed females and the red-eyed males obtained in this study were* XXY *and* XO, *respectively. Subsequent cytological analyses confirmed that they were really so. This clearly demonstrated that the gene* v *was located in the* X *chromosome of* Drosophila.

Attached-*X* Chromosome

- Discovered by L.V. Morgan in 1922 for *yellow* body in *Drosophila.*
- A female showed regular unusual inheritance of the sex-linked *yellow* body gene, *viz.*, yellow ♀ × gray ♂ → F_1: ♀♀ yellow, ♂♂ gray.
- The two *X* chromosomes in this female were fused to form a single chromosome; *confirmed cytologically.*
- Confirms that *yellow* gene is located in *X* chromosome.

21.4. ATTACHED-X CHROMOSOME IN *DROSOPHILA*

In 1922, L.V. Morgan obtained a female *Drosophila* with yellow body colour. This female, when mated with a gray body male, produced all females with yellow body and all males with gray body colour. Yellow body is recessive to normal gray body. It is determined by the gene *Y* located in the *X* chromosome. Clearly, this yellow body female was producing only unusual females and males in its progeny. In comparison, Bridges had obtained the unusual individuals only in a low frequency (one fly per 2-3 thousand flies).

L.V. Morgan reasoned that nondisjunction of the *X* chromosome was regularly taking place in all the oocytes of this female (X^yX^y). Therefore, it produced only the two unusual egg types: X^yX^y and *O* (without an *X* chromosome). Such a highly regular nondisjunction is only possible if the two *X* chromosomes behaved as a single chromosome during meiosis, that is, both shared the same centromere. In other words, the two *X* chromosomes should have fused to form a single chromosome, the ***attached-X chromosome*** (Fig. 21.4). A cytological analysis confirmed that the *X* chromosomes of this female were actually fused into one. Such an attached-*X* chromosome may be represented as X-X. X-X regularly passes as a unit to the same pole during AI, while the opposite pole does not receive any *X* chromosome.

The yellow body female with an attached-*X* (X^y-X^y) would produce only the two types of unusual eggs: X^y-X^y and *O*. The gray body male (X^YY) will produce two types of sperms, X^Y and *Y*. A union between these male and female gametes will produce the following four types of zygotes: X^y-X^yX^y, X^y-X^YY, X^YO and OY. The first combination will produce a superfemale, which normally dies. The fourth combination (OY) does not survive. The X^y-X^yY zygotes receive both their *X* chromosomes from their female parent. Therefore, they will develop into yellow-body females. The third combination (X^YO) receives its *X* chromosome from the male parent. Hence it will produce gray-body males.

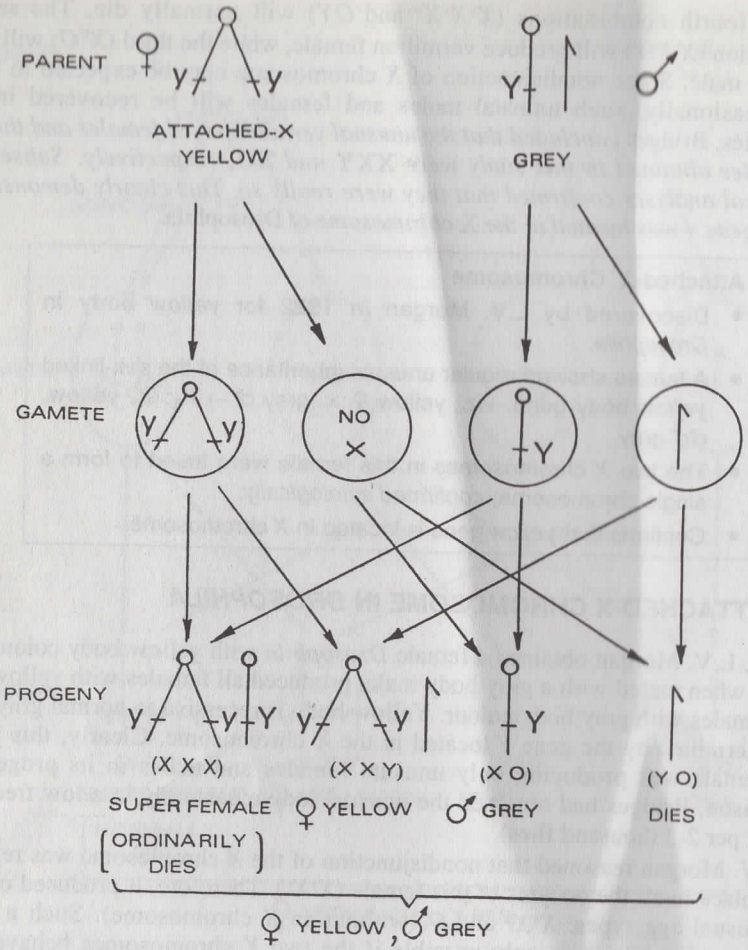


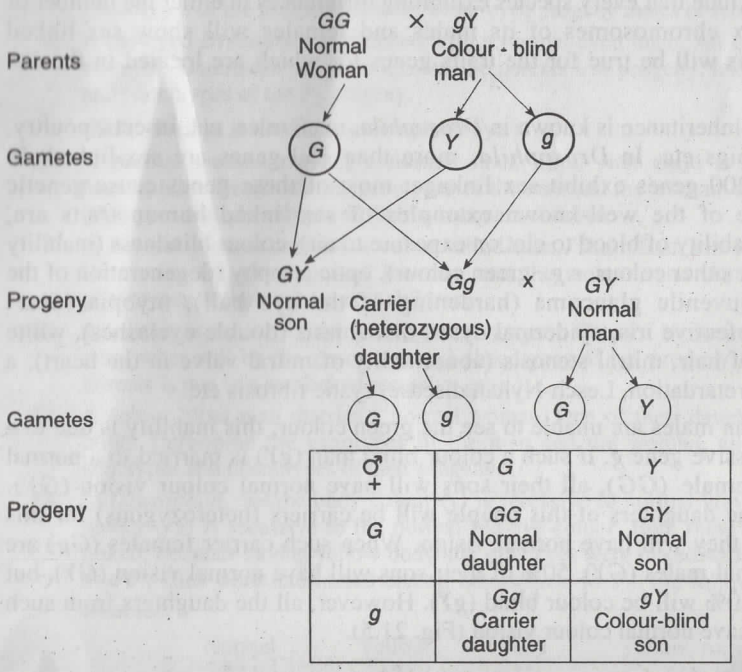
Fig. 21.4. Regular male-to-male transmission of the dominant allele of *yellow* body gene (*Y*) due to the attached-*X* in *Drosophila.*

Thus unusual transmission of the X *chromosome and, consequently, of* X*-linked genes from male parent to male progeny could be explained by the nondisjunction of* X *chromosome in the female.* There was a complete correspondence between the expectations from this hypothesis and the actual cytological data on the unusual progeny in the studies of Bridges and L.V. Morgan. These studies, therefore, provided a conclusive proof that sex-linked genes are located in the *X* chromosome of *Drosophila. Thus these findings established the chromosomal basis of heredity beyond any reasonable doubt.*

21.5. CHARACTERISTICS OF SEX-LINKED INHERITANCE

The characteristic features of inheritance of a sex-linked trait may be summarised as follows.

1. The frequency of individuals showing a recessive sex-linked trait is markedly higher in the heterogametic sex (*e.g.,* male *Drosophila* and humans, and female birds) than that in the homogametic sex (*e.g.,* female *Drosophila* and humans, and male birds).
2. Ordinarily, genes governing sex-linked traits in man and *Drosophila* are not transmitted from male parents directly to their male progeny. For example, *white* eye gene (*w*) is not transmitted from male *Drosophila* to its male progeny.
3. A male human or *Drosophila* transmits its sex linked genes to all its daughters. These daughters transmit this gene to half of their male progeny. As a result, *a sex-linked recessive gene is transmitted from a male to its female progeny and then to half the male progeny of such females. Thus a* ***sex-linked gene passes from male to female then back to male;*** *such an inheritance pattern is known as* ***criss-cross inheritance.***
4. Sex-linked genes are not present in the *Y* chromosome. Consequently, the heterogametic sex (male humans, and *Drosophila,* and female birds) is hemizygous for such genes, *i.e.,* it has only one allele of the sex-linked genes.
5. It was noted that a region of the *Y* chromosome is homologous to a region in the *X* chromosome. For example, the long arm of human *Y* chromosome is homologous to the short arm of human *X* chromosome. Similarly, one-third of the *Drosophila X* chromosome (the region close to the centromere)



♂ / ♀	G	Y
G	GG Normal daughter	GY Normal son
g	Gg Carrier daughter	gY Colour-blind son

Fig. 21.5. A colour-blind man transmits the defective gene to all his daughters. These daughters are normal, but are carriers for the gene *g*. When the daughters marry normal men, 1/2 of their sons are colour-blind. Thus the sex-linked gene *g* passes from father-to-daughter-to-grandson.

is homologous to a region of the *Y* chromosome. *Genes located in such regions of* X *chromosome do not show sex-linked inheritance.* (These genes show normal disomic inheritance pattern as exhibited by autosomal genes.)

21.6. THE CAUSE OF SEX-LINKAGE

The peculiar inheritance pattern of sex-linked traits is due to the following two reasons: (1) ***the location of a gene in the X chromosome, and*** (2) ***the absence of its allele in the Y chromosome.*** These two features were postulated by Morgan in 1910 on the basis of the inheritance pattern of *white* eye gene of *Drosophila*. Experimental evidence in support of these postulates came from the studies of Bridges and L.V. Morgan.

Sex Linkage in Man

- Over 200 genes are sex-linked (located in *X* chromosome).
- Most genes cause genetic diseases, *e.g.*, haemophilia, optic atrophy, myopia, colour blindness, cystic fibrosis, etc.
- Males show much higher frequencies of such traits than females.

21.7. SEX LINKAGE IN MAN AND OTHER ORGANISMS

It is safe to conclude that every species exhibiting differences in either the number or the kind of sex chromosomes of its males and females will show sex-linked inheritance. This will be true for the traits genes for which are located in the sex chromosome.

Sex-linked inheritance is known in *Drosophila,* man, mice, cat, insects, poultry, cattle, guinea pigs etc. In *Drosophila,* more than 150 genes are sex-linked. In humans, over 200 genes exhibit sex linkage; most of these genes cause genetic diseases. Some of the well-known examples of sex-linked human traits are, haemophilia (inability of blood to clot on exposure to air), colour blindness (inability to see one or the other colour, *e.g.*, green colour), optic atrophy (degeneration of the optic nerve), juvenile glaucoma (hardening of the eye ball), myopia (near-sightedness), defective iris, epidermal cysts, distichiasis (double eyelashes), white occipital lock of hair, mitral stenosis (abnormality of mitral valve in the heart), a form of mental retardation, Lesch-Nyhan disease, cystic fibrosis etc.

Some human males are unable to see the green colour; this inability is due to a sex-linked recessive gene *g*. If such a colour blind man (gY) is married to a normal homozygous female (GG), all their sons will have normal colour vision (GY). However, all the daughters of this couple will be carriers (heterozygous) for this gene (Gg), but they will have normal vision. When such carrier females (Gg) are married to normal males (GY), 50% of their sons will have normal vision (GY), but the remaining 50% will be colour blind (gY). However, all the daughters from such marriages will have normal colour vision (Fig. 21.5).

QUESTIONS

1. Explain sex linkage with the help of a suitable example. List the characteristic features and explain the reasons for this pattern of inheritance.

2. Explain criss-cross inheritance giving a suitable example. Discuss the reasons for this inheritance pattern.
3. Explain the reasons for the following:
   (*i*) Sex linkage.
   (*ii*) Sex-linked traits are more frequent in men than in women.
   (*iii*) A sex-linked gene is not transmitted from father to his son.
   (*iv*) In *Drosophila*, red-eyed males are obtained in very low frequency from matings between white-eyed females and red-eyed males.
   (*v*) All male progeny of some white-eyed females mated to red-eyed males are red-eyed.
4. Write short notes on the following:
   (*i*) sex linkage,
   (*ii*) criss-cross inheritance,
   (*iii*) sex linkage in man,
   (*iv*) non-disjunction of *X* chromosome,
   (*v*) attached—*X* chromosome.
5. A white-eyed female *Drosophila* is mated with a red-eyed male. Determine the genotypes and phenotypes of its progeny. (Remember, white eye is recessive to normal dull red eye).
6. A red-eyed female *Drosophila* is mated with a red eyed male. About 25% of the progeny are white-eyed. Determine the genotypes of the parents, and genotypes and phenotypes of the progeny. Are the white-eyed progeny males or females?
7. A bar-eyed *Drosophila* was mated with a normal-eyed male. All the progeny had bar eyes. Determine the genotypes of the parents and progeny, and the genotypes and phenotypes of the F_2 progeny.
   (**Note:** *bar* eye is dominant over normal eye.)
8. A normal woman marries a normal man. Their first child is a male and is haemophilic. Determine the genotypes of the man and the woman, and the possible genotypes and phenotypes of their subsequent children.
9. A normal woman marries a colour-blind man. Their first son is colour-blind. Determine the genotypes of the man and the woman, and the genotypes and phenotypes of their subsequent children.
10. A women is carrier for the gene governing cystic fibrosis. If she marries a normal man, what will be the genotypes and phenotypes of her progeny ? (**Note:** cystic fibrosis is due to a sex-linked recessive gene.)
11. A colour-blind man marries a normal woman. One of their daughters is colour-blind. Determine the genotypes of the man and the woman, and the probable genotypes and phenotypes of their progeny.
12. In *Drosophila, yellow* body colour (*y*) is recessive to normal gray, and *ruby* eye colour (*rb*) is recessive to the normal red eye colour. Both these genes are sex-linked. Females having yellow body and ruby eyes were mated with normal males. The F_1 flies were intermated among themselves. In F_2, the following data were obtained.

Normal	Yellow	ruby	yellow, ruby
577	47	51	565

   (*i*) Depict this cross using the given gene symbols,
   (*ii*) estimate the frequency of crossing over between genes *y* and *rb*, and
   (*iii*) prepare there linkage map.

13. In *Drosophila, carnation* eye colour (*car*) is recessive to the normal red, and *bar* eye shape (*B*) is dominant to the normal oval. Both the genes *car* and *B* are present in the *X* chromosome. Female flies heterozygous for *carnation* and *bar* genes were mated with carnation, normal male flies. Only the male progeny from this cross were scored to yield the following data.

Normal	Carnation	Bar	Carnation, bar
40	652	645	43

(*i*) Depict this cross using the given gene symbols,
(*ii*) estimate the frequency of recombination between *car* and *B*, and
(*iii*) prepare their linkage map,
(*iv*) If males homozygous for the genes *car* and *b* are mated with carnation, normal females, what type of progeny will be obtained?
(*v*) If the heterozygous females were mated with normal, bar males, will the results have been any different?

14. In *Drosophila, cut* wings (*ct*) is recessive to normal wings, and *vermilion* eye colour (*v*) is recessive to red. The two genes are present in the *X* chromosome 13.0 crossover units away from each other. Females with the genotype *Ct v/ct V* were mated with normal males. Depict this cross, and estimate the frequencies of the different types of progeny obtained.

15. If *Ct v/ct V* females were mated with cut, vermilion males, what will be the per cent frequencies of the different types of progeny obtained. If the total number of the progeny obtained was 1000, give the expected numbers of the different types of progeny.

16. In *Drosophila white* eyes (*w*) are recessive to red. Gene *w* is located about 18 crossover units away from the gene *ct*. *W Ct/w ct* females were mated with normal males. Determine the per cent frequencies of the different types of progeny obtained from this cross. If the total number of progeny were 2,200, give the expected frequencies of the different phenotypic classes.

22
Mutation

22.1. INTRODUCTION

The replication and distribution of genetic material is extremely precise so that the genetic information is usually passed on from one generation to the next without alternation. But occasionally, errors do occur both in replication and distribution of the genetic material. This gives rise to sudden heritable changes in the characters of organisms; such alterations are called ***mutation***. The individuals showing these heritable changes are known as ***mutants.*** But an individual exhibiting only an altered phenotype is referred to as a ***variant*** till it is established that the change is heritable. The allele of a gene producing the changed phenotype is called ***mutant allele***. Clearly, ***a mutation is a change in phenotype, which is sudden, heritable and is not produced due to segregation or recombination.***

Mutation
- A sudden, heritable change in phenotype.
- Types:
  - Forward, reverse
  - Spontaneous, induced
  - Dominant, recessive, codominant, incompletely dominant
  - Somatic / bud, germinal
  - Lethal, sublethal, subvital, vital, supervital
  - Visible, biochemical
  - Macro–, micromutation
  - Gene/point, chromosomal, cytoplasmic/plasmagene
  - Base substitution, deletion, addition
  - Missense, nonsense, frame-shift

A mutation may arise due to a change in the base sequence of a gne; such mutations are called ***gene mutations*** or ***point mutations***. Similarly, mutations resulting from changes in the plasma-genes are termed as ***plasma gene mutation***. But changes in chromosome number and structure also produce heritable changes in phenotype (Chapter 5); these are termed as ***chromosomal mutations.*** In most eukaryotes, it is generally not possible to detect at least some types of structural chromosome changes (*e.g.*, small deletions and duplications), while numerical chromosome changes are readily observable. As a result, heritable changes produced by many structural chromosome changes would be classified as gene mutations in most eukaryotes. The type of change producing a mutation can be determined with a great degree of precision in prokaryotes, and most of our knowledge about mutations is based on studies in bacteria and viruses.

In this chapter, the *term* ***mutation*** *has been used as a synonym of gene mutation.* ***Mutation is the ultimate source of all the genetic variation existing in any organism.*** *Without mutation all the genes would be present in only one form (the wild type alleles). In the absence of their alternative forms or alleles, evolution and genetic studies would be impossible.*

22.2. HISTORICAL

The term mutation was introduced by Hugo de Vries in 1900 to describe sudden heritable changes in *Oenothera.* But the earliest record of a sudden, heritable change (mutation) dates back to 1791. An English farmer named Seth Wright discovered a male lamb with unusually short legs. This lamb served as the source of dominant short-legged trait for the development of the 'Ancon' breed of sheep.

Systematic studies on mutations began in 1910 with the discovery of the *white* eye mutant of *Drosophila* by Morgan. Soon geneticists began to search for ways to induce mutations. H.J. Muller discovered the mutagenic action of X-rays in 1927 in *Drosophila*. He was awarded the Nobel Prize for Physiology and Medicine in 1946. Two years later, in 1929, Stadler described the mutagenic effects of X-rays in barley. Subsequently, mutagenic effects of mustard gas and some other chemical compounds were discovered by Auerbach and Robson in 1946. The first plant breeding programme designed to exploit induced mutations (***mutation breeding***) was initiated in 1929 in Sweden, Germany and U.S.S.R. In India, the first mutation breeding programme began in the early 1930's.

22.3. CHARACTERISTICS OF MUTATIONS

The available evidence shows that ***there is no difference in the nature of spontaneous and induced mutations.*** The main characteristics of mutations may be summarised as follows.

1. Generally, *mutant alleles are recessive* to their wild type or normal alleles. But some mutations are dominant, *e.g., Notch* wing in *Drosophila.*
2. *Mutations are random* events in terms of the time of their occurrence and the gene in which they occur. Thus mutation of a specific gene is almost entirely a matter of chance.
3. *Most mutations have harmful effects*, but some mutations are beneficial.
4. *Mutations are recurrent.* That is, the same mutation may be expected to occur in different individuals of a given generation or those belonging to different generations.
5. *Mutations occur at very low frequencies in nature*; such mutations are called ***spontaneous mutations.*** The rate of spontaneous mutations for most of the genes is very low, ranging between 10^{-7} and 10^{-4}.
6. The rate of spontaneous mutations varies considerably from one gene to another (Table 22.1).
7. Some genes show exceptionally high rates of spontaneous mutations and are known as ***mutable genes.***
8. Some genes increase the spontaneous mutation rates of some other genes of the genome; such genes are called ***mutator genes.***
9. Certain other genes suppress the mutation rate of some other genes of the genome. Such genes are termed as ***antimutator genes.***

TABLE 22.1

Spontaneous mutation rates for some genes in different organisms. In most of the cases, the rates are for forward mutation

Organism	*Character affected and direction of mutation*	*Mutation rate per 10^6 gametes or cells*
Escherichia coli	*ara*$^+$ → *ara*$^-$ (arabinose auxotroph)	2
	leu → *leu*$^+$ (leucine prototroph)	0.0007
Drosophila melanogaster	*y*$^+$ → *y* (yellow body)	120
	bw$^+$ → *bw* (brown eye)	30
	ey$^+$ → *ey* (eyeless)	60
Zea mays	*wx*$^+$ → *wx* (waxy starch)	0
	c$^+$ → *c* (colourless)	2.3
	su$^+$ → *su* (sugary endosperm)	2.4
	pr$^+$ → *pr* (purple)	11
Mouse	*c*$^+$ → *c* (albino)	10.2
Human	Normal → epiloia	4 – 23
	Normal → Huntington's chorea	5
	Normal → achondroplastic dwarfness	42 – 143

10. Mutations producing the same phenotypic effect may be located at different sites within the same gene.
11. Some sites within genes show very high rates of mutation, while some other sites show little or no mutation. The highly mutable sites within a gene are commonly known as ***hot spots.***
12. Many agents, both physical and chemical, increase the frequency of mutations by several-fold. These agents are known as ***mutagens*** (Table 22.2).
13. The rate of induced mutations varies considerably from one gene to the other.
14. Mutations can occur in any tissue/cell (either somatic or germinal) of an organism.
15. Mutational events can occur during any developmental stage of organisms. But treatments with mutagens during certain stages of cell division (*e.g., S* phase) may yield mutations at a rate higher than those obtained by treatments during other phases of the cell cycle.
16. The frequency of induced mutations is markedly affected by environmental conditions prevailing just prior to, during and even after the mutagen treatment.
17. *Most mutant alleles are pleiotropic*. That is, a single mutant allele affects more than one character.
18. In case of many genes, mutant individuals show markedly increased variation in their phenotypes as compared to the normal individuals.
19. Mutations occur in both ***forward*** (from wild type allele to mutant allele) and ***reverse*** (from mutant allele to wild type allele) directions.

20. Generally, *the rates of forward mutations are much higher than those for reverse mutations.*
21. *Some mutant alleles do not mutate back. That is, they do not show reverse mutation. Such mutant alleles are believed to be produced by deletions.*
22. *Several different mutations may occur in a single gene giving rise to multiple allelic series at such loci.*

TABLE 22.2

Some of the commonly used physical and chemical mutagenic agents. (Agents most commonly used in plants are marked with an asterisk.)

Class	*Mutagen*	*Remarks*
	Physical Mutagens	
Ionizing Radiations		
Particulate radiations	alpha-rays (α-rays)	2 protrons + 2 neutrons, densely ionizing (DI)
	Beta-rays (β-rays)	High energy electrons; sparsely ionizing (SI)
	Fast neutrons*	DI
	Thermal neutrons*	DI
Nonparticulate radiations	X-rays*	0.01 - 1nm**; SI
	Gamma-rays (γ–rays)*	0.001 - 1nm; SI
Nonionizing Radiations	Ultraviolet rays (UV rays)	10 - 390 nm
	Chemical Mutagens	
Alkylating Agents	Mustard gas or sulphur mustard, Nitrogen mustard, Ethylmethane sulphonate (EMS)*, Methylmethane sul-phonate (MMS), Ethylethane sulphonate (EES), N-Methyl-N'-nitro-N'-nitrosogunani-dine (NTG)	
Base Analogues	5-Bromouracil (5-BU) 2-Aminopurine (2-AP)	
Acridine Dyes	Acriflavin, Proflavin, Acridine orange	
Deamination Agents	Ethidium bromide	Potent mutagen for plasmagenes
	Nitrous acid (HNO_2)	
Other Chemical Mutagens	Hydroxylamine (HA) Sodium azide*	
	DNA sequences***	

* Mutagens commonly used in plants.

** nm = nanometer = 10^{-9} m = 10^{-6} mm = 10Å (angstrom).

*** Oligonucleotides, transposons and insertion sequences are used for mutagenesis *in vitro* and in prokaryotes. T-DNA is frequently used in plants.

22.4. CLASSIFICATION OF MUTATIONS

There are several ways in which mutations can be classified; some of the common ones are listed below.

22.4.1. Direction of Mutation

Mutations are grouped into the following two classes on the basis of direction of the mutational event: (1) forward and (2) reverse mutation. A mutation from wild type allele to a mutant allele is called ***forward mutation.*** But mutation from a mutant allele to the normal one is known as ***reverse mutation.***

22.4.2. Cause of Mutation

Mutations are classified as (1) spontaneous or (2) induced depending on the cause of their origin. ***Spontaneous mutations*** occur naturally without any apparent cause. But ***induced mutations*** originate due to a treatment with some physical or chemical agent.

22.4.3. Dominance Relationship

Mutant alleles may be (1) dominant (***dominant mutation***), (2) most often recessive (***recessive mutation***), (3) occasionally codominant (***codominant mutation***), or (4) partially dominant (***incompletely dominant mutation***).

22.4.4. Tissue of Origin

Mutations may be classified as (1) somatic and (2) germinal on the basis of tissue in which they occur. A mutation occurring in a somatic cell is called ***somatic mutation***. When a somatic mutation occurs in an axillary bud, it is termed as ***bud mutation***. ***Germinal mutations,*** on the other hand, occur in reproductive tissues. In a sexually reproducing species, only germinal mutations are transmitted to the next generation.

22.4.5. Effect on Survival

Many mutations have drastic effect on survival; they are classified as (1) lethal, (2) sublethal, (3) subvital, (4) vital, and (5) supervital. ***Lethal mutations*** kill all the individuals having them in the appropriate genotype, ***sublethal mutations*** kill most of the individuals, while ***subvital mutations*** kill some of the individuals. ***Vital mutations*** do not affect survival, while ***supervital mutations*** increase the survival.

22.4.6. Type of Trait Affected

Mutations may be grouped into the following categories: (1) visible or morphological mutations, and (2) biochemical mutations. A mutation that alters a morphological feature of the organism so that the presence of the mutant allele is visually detected, is called a ***visible mutation***. On the other hand, those mutations, which prevent the production of a biochemical by the organism are referred to as ***biochemical mutations***.

22.4.7. Quantum of Morphological Effect Produced

Mutations may be classified as (1) macro-mutations and (2) micro-mutations depending on the quantum of phenotypic change produced by them. ***Macromutations*** produce large enough changes in the phenotype that can be detected without any confusion, *e.g.,* short-legged mutation in sheep. In general, mutations in qualitative traits are of this type. Mutations in quantitative characters generally produce small morphological effects, which are not easily detected. Special techniques are, therefore, required for their detection. Such mutations are called ***micromutations.***

22.4.8. Cytological Basis

Mutations are called (1) chromosomal, (2) gene, or (3) cytoplasmic depending on their cytological basis. ***Chromosomal mutations*** are due to detectable changes in either chromosome number or structure, *e.g., bar* eye (*B*) mutation in *Drosophila*. But mutations produced by alternations in the base sequences of genes are known as ***gene mutations*** (or ***point mutations***). Both these types of mutations involve chromosomes and nuclear genes. Another class of mutations, ***cytoplasmic*** or ***plasmagene mutations***, are associated with changes in mitochondria or chloroplast DNA. These mutations show typical cytoplasmic inheritance, *e.g.*, cytoplasmic male sterility (cms) in maize, sorghum, pearl millet, etc.

22.4.9. Molecular Basis

Gene mutations can be subdivided into the following three classes on the basis of the molecular change: (1) base substitution, (2) deletion, and (3) addition. When a single base in a DNA molecule (gene) is replaced by another base, it is called ***base substitution.*** In a ***deletion mutation***, one or more bases are deleted or lost from a gene. On the other hand, insertion of one or more bases in a gene is referred to as ***addition.***

22.4.10. Type of Amino Acid Replacement

Mutations are called (1) missense, (2) nonsense, or (3) frame-shift mutations depending on the type of amino acid replacements they produce. ***Missense mutations*** cause the replacement of a single amino acid of the respective polypeptides by another amino acid. They are produced by base substitutions. A ***nonsense mutation*** produces a nonsense codon. Such a codon does not code for any amino acid. Consequently, nonsense codons terminate the growth of polypeptide chains. In many mutations, all the amino acids of the polypeptide chain located beyond the site of mutation are substituted. Such mutations are called ***frame-shift mutations.*** *They are produced due to deletion or addition of bases.*

Spontaneous Mutation

- Occur without any treatment.
- Frequency $10^{-4} - 10^{-7}$ per gene / generation.
- *Produced due to*:
  - (*i*) Errors by DNA polymerase.
  - (*ii*) Transposable elements.
  - (*iii*) DNA methylatation, followed by deamination.
  - (*iv*) Environmental factors.

22.5. SPONTANEOUS MUTATION

Spontaneous mutations occur naturally without any apparent or known cause. The possible sources of origin of spontaneous mutations are as follows: (1) errors during DNA replication, (2) mutagenic effects of the natural environments of organisms, (3) transposons and insertion sequences, and (4) methylation, followed by spontaneous deamination of DNA bases, especially cytosine.

The contribution of environment to spontaneous mutations is not clearly understood nor can be unequivocally demonstrated. Some of the solar radiations are surely mutagenic and are likely contributors to spontaneous mutations. For example,

ultraviolet (UV) rays of sunlight are highly mutagenic. But they are unable to penetrate more than a couple of cell layers in animals and higher plants. In addition, many naturally occurring compounds may be mutagenic. Both eukaryotic and prokaryotic genomes contain some transposable elements, *e.g.*, insertion sequences (IS), transposons, etc. These elements integrate at specific sites in the genome and produce gene mutations.

Another likely source of spontaneous mutations is the error committed by DNA polymerase during DNA replication. It is estimated that DNA polymerase incorporates wrong bases with a frequency of 10^{-5}. Proof-reading by DNA polymerase during replication is estimated to reduce this frequency to 10^{-10}. If an average gene is assumed to be of 1,000 bp, the frequency of incorporation of wrong bases will be 10^{-7} bases per gene per round of replication. Many of these errors will not produce mutations due to their location in noncritical sites of the polypeptide or even due to subsequent repair.

A major cause of spontaneous mutations in *E. coli* is the modification of DNA after synthesis, *e.g.*, methylation of cytosine at position 5 by a methylase enzyme. An appreciable proportion of the 5-methylcytosine undergoes spontaneous deamination, which yields thymine. Replication of such a molecule leads to G.C → A.T transition at such sites where 5-methylcytosine had undergone deamination.

The rate of spontaneous mutations is very low (Table 22.1). It generally ranges between 10^{-5}–10^{-7}/gene/generation for prokaryotes and $10^{-4} - 10^{-6}$/ gene/generation for eukaryotes. The rates of reverse mutations have been estimated in prokaryotes. These rates are generally much lower than those for forward mutations.

In general, spontaneous mutation rates per generation are comparatively higher in eukaryotes than in prokaryotes. But some eukaryotic genes are highly stable and mutate at a very low rate, *e.g.*, the *waxy* locus in maize. Estimated mutation rates in man appear to be much higher than those for other eukaryotes. Finally, different genes of a single species may show markedly different mutation rates. For example, the rates of forward mutations to *yellow* body colour and to *brown* eye colour in *Drosophila* are estimated as 1.2×10^{-4} and 3×10^{-5}, respectively.

Induced Mutations

- Induced by treatment with a mutagen:
  - (*i*) Physical mutagenes.
  - (*ii*) Chemical mutagens.

22.6. INDUCED MUTATIONS

Mutations produced due to the treatment with either a chemical or a physical agent are called ***induced mutations.*** The agents capable of inducing mutations are known as ***mutagens.*** The capacity of an agent for inducing mutations is termed as ***mutagenic property***. Induced mutations are useful in two different ways: (1) in genetic and biochemical studies, and (2) in crop improvement. The process of inducing mutations through treatment with a mutagen is known as ***mutagenesis.*** The exploitation of induced mutations for crop improvement is called ***mutation breeding.***

Treatments with various mutagens increase the mutation rate by several orders of magnitude. Muller observed upto 150-fold increase in mutation rate following

irradiation of *Drosophila* with X-rays. Some potent chemical mutagens may produce mutations at the rate of more than 1% per gene per generation in bacteria and bacteriophages. This represents a more than 10,000-fold increase in the mutation rate.

22.7. MUTAGENS

Agents capable of inducing mutations are called ***mutagens.*** The different mutagenic agents may be classified into two broad group: (1) physical mutagens and (2) chemical mutagens.

UV Rays

- Nonionizing.
- Poorly penetrating.
- Cause atom excitation.
- *Direct effects* :
  - (*i*) Pyrimidine dimer formation.
  - (*ii*) Pyrimidine hydration.
- Indirect effects: due to free radicals and organic peroxides
- Wavelength 10–390 nm.
- 254 nm most mutagenic; greatest absorption by DNA bases.
- White light suppresses UV action.

22.7.1. Physical Mutagens

The different types of radiations having mutagenic properties are known as ***physical mutagenes.*** *These radiations are high energy radiations.* They may be either nonparticulate, *e.g.,* UV rays, X-rays, gamma-rays and cosmic rays, or particulate, *e.g.,* α-rays, β-rays, etc. Radiations are grouped into the following two classes depending on the kind of effect they have on the atoms in their path: (1) ionizing and (2) nonionizing radiations (Table 22.2).

22.7.1.1. Noninonizing Radiation

Ultraviolet rays (present in solar radiation or produced by mercury-vapour lamps or tubes) are the only nonionizing radiation with mutagenic properties. Their wavelength ranges from 100-3,900 Å (10-390 nm). UV rays are specifically absorbed by purines and pyrimidines present in DNA. The maximum absorption of UV rays by DNA occurs at the wavelength of 254 nm. This is also the most mutagenic wavelength of UV. UV is relatively low energy radiation. It has a very poor penetrating power; it can penetrate only the surface layer of cells in multicellular organisms. But it is highly mutagenic in microorganisms.

The mutagenic action of UV rays is the consequence of both its direct and indirect effects on DNA. The ***direct effect*** of UV on DNA is the formation of (1) pyrimidine dimers and (2) pyrimidine hydrates. A photon of UV light does not have sufficient energy to cause ionization, *i.e.,* removal of an electron (e^-) from an atom. It only raises the energy level of an e^- so that it moves into an outer orbit of the same atom. This is called ***excitation.***

Primidines, especially thymine, having excited atoms are highly reactive. They readily form ***pyrimidine dimers*** (Fig. 22.1). Thymine dimers are the most commonly

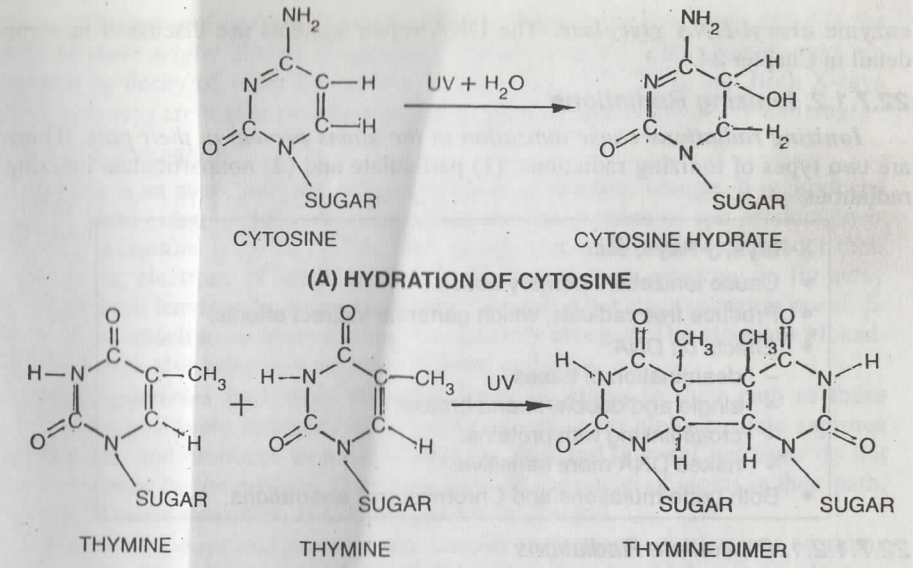


Fig. 22.1. Effects of UV-rays on pyrimidines. (A) Hydration of a cytosine molecule due to the addition of an –H at 4 and an –OH at 5 positions; the double bond between 4 and 5 carbons is disrupted. (B) the formation of a thymine dimer due to covalent bond formation between the two 4 carbons and the two 5 carbons of the two thymine molecules lying next to each other in the same strand of a DNA molecule.

formed, but cytosine dimers and thymine-cytosine dimers are also produced. Dimer formation involves two pyrimidine molecules located next to each other in the same DNA strand. DNA polymerase is unable to catalyse the replication of a DNA molecule in the region containing a pyrimidine dimer. Another direct effect of UV on DNA is the ***hydration of pyrimidines*** (Fig. 22.1). Generally cytosine is hydrated, but hydration of thymine also occurs. *A major portion of the mutagenic action of UV rays is believed to be due to pyrimidine dimer formation.*

UV rays also have an indirect mutagenic effect. When unirradiated bacterial cells are placed onto a UV-irradiated medium, mutations are induced in these cells. The *indirect effect is suggested to be due to organic peroxides and free radicals produced in the medium due to UV-irradiation. But the effect of UV on medium is short-lived, lasting for only a few hours.*

Both eukaryotes and prokaryotes possess extensive ***DNA repair enzyme systems*** for repairing the damage caused to DNA by UV rays and, possibly, other radiations. Exposure of organisms to visible light, particularly to blue light, after UV irradiation greatly reduces the yield of mutations. Bacterial cells are readily killed (inactivated) on exposure to UV rays. The frequency of inactivated cells drops sharply if UV-irradiated cells are exposed to visible light. Both the above phenomena are called ***photoreactivation.*** *This mechanism breaks the bonds between thymine dimers.* A second repair mechanism, called ***excision-repair,*** *involves the removal of thymine dimers.* The latter mechanism does not require blue light. The third mechanism of DNA repair in *E. coli* involves ***DNA replication and recombination*** between the two 'sister' molecules thus produced. A fourth repair mechanism is based on the

enzyme ***uracyl-DNA glycolase.*** The DNA repair systems are discussed in some detail in Chapter 24.

22.7.1.2. Ionizing Radiations

Ionizing radiations *cause ionization in the atoms present in their path.* There are two types of ionizing radiations: (1) particulate and (2) nonparticulate ionizing radiations.

X-Rays, γ-Rays, etc.

- Cause ionization: primary action.
- Produce free radicals, which generate indirect effects.
- *Effects on DNA*:
  - deamination of bases.
  - single and double-strand breaks.
  - cross-linking with proteins.
  - naked DNA more sensitive.
- Both gene mutations and Chromosomal aberrations.

22.7.1.2.1. Particulate Radiations

These consist of high energy atomic particles produced by radioactive decay. ***Beta-rays*** are high energy electrons produced by radioactive decay of ^{3}H, ^{32}P, ^{35}S, etc.; they are negatively charged. ***Alpha-rays*** are composed of alpha-particles having two protons and two neutrons each. Alpha-particles have double positive charge. ***Fast neutrons*** are produced in cyclotrons or atomic reactors due to radioactive decay of heavier elements. The velocity of fast neutrons is reduced by graphite or heavy water to generate ***thermal*** or ***slow neutrons***. Both the types of neutrons are uncharged particles, and are highly penetrating in biological tissues. Alpha-particles and the two types of neutrons are densely ionizing radiations.

22.7.1.2.2. Nonparticulate Ionizing Radiations

These radiations are represented by X-rays and gamma-rays. They are high energy radiations composed of photons. *The physical properties and biological*

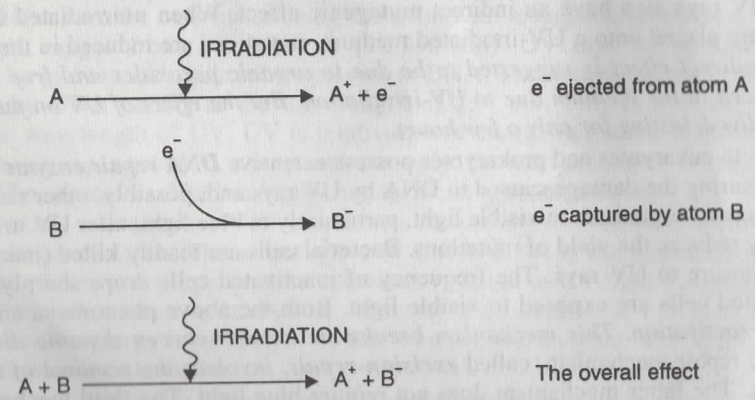


Fig. 22.2. The production of ion-pairs by high energy radiations

effects of ***X-rays*** *and* ***gamma-rays*** *are similar. These radiations only differ in the source of their origin*: X-rays are produced by X-ray tubes, while gamma-rays are produced by decay of some radioactive isotopes, *e.g.*, ^{14}C, ^{60}Co, etc. Both X-rays and gamma-rays are highly penetrating in biological tissues and sparsely ionizing.

22.7.1.2.3. Ionization

An ***ion*** is an atom carrying either a positive or negative charge. It is produced when an atom either gains or loses an electron. The process of ion production is known as ***ionization*** (Fig. 22.2). The high energy electrons of ***β-rays*** transfer their energy to the electrons of atoms in their path causing these electrons to fly away from their orbit leaving the atoms positively charged. After their energy is spent, β-ray electrons attach to an atom making it negatively charged. The electrons kicked-away by β-rays also behave in the same manner as β-rays.

Alpha-particles pull away electrons from the atoms in their path as these particles are positively charged. After losing energy, each alpha-particle captures two electrons and produces an atom of helium. ***Fast*** and ***thermal neutrons*** do not cause ionization by themselves. They may kick out the nuclei of atoms in their path; these nuclei cause ionization as they are positively charged.

Photons of ***X-rays*** and ***gamma-rays*** transfer their energy to electrons of atoms in their path. The elctron is kicked off from its orbit as a high energy electron causing ionization.

22.7.1.3. Genetic Effects

The *chemical and biological (including genetic) effects of radiations are produced primarily due to ionization. The genetic effects of radiations result from their effect on DNA.* The genetic effects include change in a base (*e.g.*, deaminution), loss of a base, disruption of hydrogen bonds between complementary strands of DNA, single and double-strand breaks in DNA, and cross-linkage of chromosomal proteins. *In general, pyrimidines are more radiosensitive than purines. The DNA present in chromatin is much more protected than naked DNA.*

The genetic effects of radiations may be (1) direct or (2) indirect. The ***direct effects*** of radiations are produced due to ionization directly in the DNA molecules. But their ***indirect effects*** are produced through ionizations in molecules other than DNA. This effect is believed to be mediated by free radical formation (Fig. 22.3). ***Free radicals*** are electrically neutral molecules with an unpaired electron in their outer orbit. Free radicals transfer their energy to other molecules making them highly reactive. Free radicals also interact with water and other molecules to produce H_2O_2 (Fig. 22.3) and organic peroxides, which are highly reactive. *They are believed to be responsible for a good portion of the biological and genetic effect of radiations. But a major portion of the genetic effects of ionizing radiations may be due to their direct effects on DNA.*

The dose of X-rays and most other ionizing radiations is measured as Roentgen units (*R* units; pronounced as 'runtgen'). *One* ***Roentgen*** is that dose of radiation, which produces one electrostatic unit of charge in one cm^3 volume under a standard set of conditions. In general, *the frequency of mutations shows a linear relationship with the dose of radiations irrespective of the intensity of irradiation.*

In many studies, frequencies of two-break chromosomal aberrations were found to be proportional to square of the dose of radiation. Chromosome breakage and

resulting structural aberrations are generally regarded as associated with gene mutations.

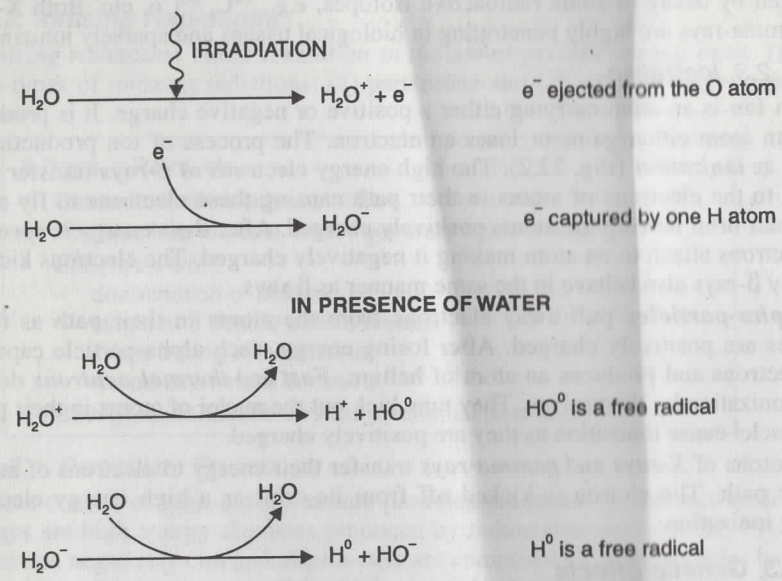


Fig. 22.3. The production of free radicals by high energy radiations; water molecules are used as an example.

Effects of irradiation are cumulative. Therefore, prolonged irradiation at low intensities (***chronic irradiation***) would produce the same amount of mutations as that produced by the same dose of irradiation delivered in a short period at a high intensity (***acute irradiation***). *Some organisms,* e.g., *mouse, apparently have an efficient DNA repair mechanism, which reduces the mutagenic effects of radiations.*

The efficiency of radiations in inducing mutations is affected by a number of factors, *e.g.,* (1) the stage of cell division, (2) O_2 concentration, (3) temperature, (4) the plant or animal species, (5) moisture content of seeds, etc. Some stages of cell division are more sensitive to radiations than others, *e.g.,* the highest number of chromosome fragments/cell was recovered when *Trillium* PMC's were irradiated (50 *R* X-rays) during diplotene, diakinesis or metaphase I.

Base Analogues

- *5-BU* : A·T → G·C transition; G·C → A·T also.
- *2-AP* : A·T → G·C.
- 5-BU normally pairs with A and occasionally with G.

22.7.2. Chemical Mutagens

A large number of chemicals is reported to be mutagenic. Some of the important chemical mutagens are listed in Table 22.2. These chemicals can be divided into the following five main classes: (1) base analogues, (2) alkylating agents, (3) acridine dyes, (4) deamination agents, and (5) other mutagenic chemicals.

22.7.2.1. Base Analogues

Base analogues *are very similar to the bases normally found in DNA. As a result, they are readily incorporated in DNA during replication.* Some of the base analogues are 5-bromouracil (5-BU), 5-chlorouracil (5-CU), 5-iodouracil (5-IU), 2-aminopurine (2-AP) and 2, 6-diaminopurine. *Of these 5-BU and 2-AP are the most commonly used base analogues.*

The mechanism of mutagenic action of 5-BU is more clearly known than that of other bases. *5-BU normally behaves like thymine and pairs with adenine. But sometimes, 5-BU pairs with guanine in place of adenine. Thus during DNA replication, 5-BU may become incorporated in place of cytosine. At the next round of DNA replication, it will normally pair with adenine. This will finally produce an A·T base pair in the place of a G·C pair. Alternatively, 5-BU may be incorporated in its normal place, i.e., in place of a thymine. But during a subsequent DNA replication, it may pair with guanine in place of adenine. This will finally produce a G·C base pair in the place on an A·T pair* (Fig. 22.4).

Alkylating Agents

- Transfer – CH_3 or $–CH_2–CH_3$ group to:
  - (*i*) PO_4 → breakage of DNA strand
  - (*ii*) 6–Oxygen or 7–nitrogen: G·C→ A·T substitution.
- Cross–linking of DNA.
- Depurination.
- Errors in DNA replication.

Thus 5-BU produces A·T → G·C as well as G·C → A·T base substitutions. Clearly, 5-BU is capable of inducing reversion of the mutations it produces. 2-AP is believed to induce mutations much in the same way as 5-BU. But it specifically induces A·T → G·C base substitutions.

22.7.2.2. Alkylating Agents

Several chemicals have reactive alkyl (e.g., methyl, $–CH_3$; and ethyl,$–CH_2–CH_3$) groups. They transfer these groups to DNA bases and to phosphate groups of DNA. These chemicals are called ***alkylating agents.*** (1) *The transfer of alkyl groups to the phosphate residue of DNA* produces phosphate triesters. This may interfere with DNA replication as well as cause breakage in the sugar-phosphate backbone.

(2) Alkyl groups are readily transferred to the 6-oxygen and 7-nitrogen of DNA bases. Of these, 7-ethyl guanine is believed to base-pair with thymine in place of cytosine. This ultimately leads to G·C → A·T base substitution.

(3) During the repair of alkylated bases, errors may occasionally occur producing base substitutions or even frame-shift mutations.

(4) Some alkylating agents cross-link DNA strands and/or DNA molecules. Such cross-links prevent DNA replication, and induce gene mutations and chromosome breakage resulting in structural chromosomal aberrations.

(5) 7-ethyl guanine often dissociates from deoxyribose and is removed from the DNA strand (***depuriation***). This may lead to the production of mutations.

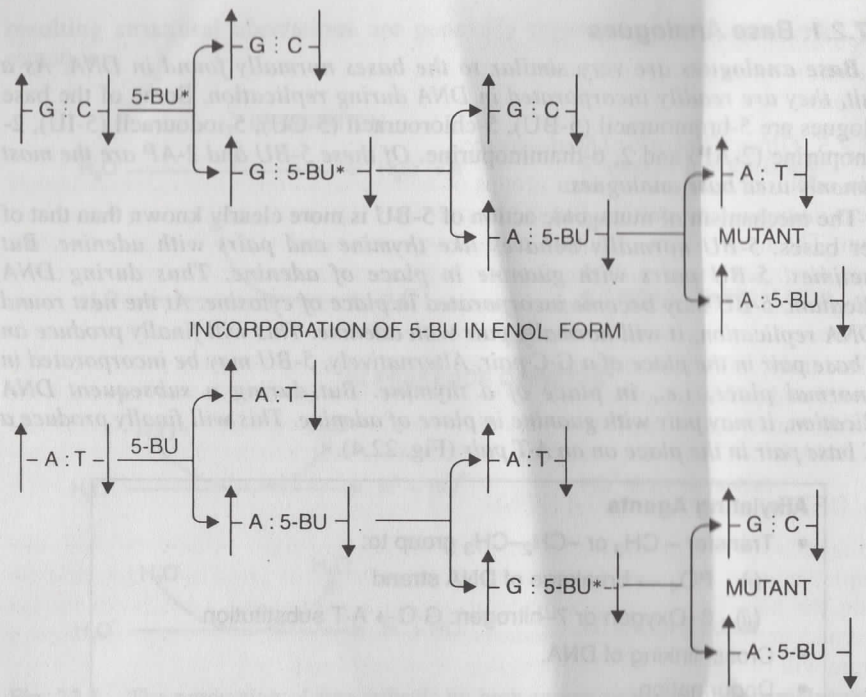


Fig. 22.4. The production of A:T → G:C and G:C → A:T transitions due to the tautometric shift in 5-BU at the time of its incorporation or after its incorporation in DNA. The enol form of 5-BU is marked by an asterisk (*e.g.* 5-BU*).

Other Chemicals

- *Nitrous acid*: Oxidative deamination of *C* and *A*; G·C → A·T; A·T → G·C.
- *Acridine dyes*: Intercalation; addition / deletion.
- *Hydroxylamine* (NH_2OH): G·C → T·A substitution.

22.7.2.3. Deamination Agents

Nitrous acid is a potent mutagen, which acts both on replicating and nonreplicating DNA. It leads to ***oxidative deamination*** of cytosine, adenine and guanine, which yields uracil, hypoxanthine and xanthine, respectively. Uracil base-pairs with adenine leading to G·C → A·T substitution. Similarly, hypoxanthine pairs with cytosine producing A·T → G·C substitution. But xanthine pairs with cytosine (the same as guanine), hence deamination of guanine does not yield mutation.

22.7.2.4. Acridine Dyes

Acridine dyes, *e.g.*, acriflavin, proflavin, etc., are very powerful mutagens, particularly in prokaryotes. *Since acridines are positively charged, they insert themselves between two base-pairs of DNA. This is known as* ***intercalation.*** *Intercalation increases the rigidity of DNA double helix as well as disturbs its*

configuration. Replication of intercalated DNA molecules often results in addition or deletion of one to few (usually one) base pairs, which produces frame-shift mutations.

22.7.2.5. Other Chemical Mutagenes

In addition to the above four groups of mutagens, a number of other chemicals possess mutagenic properties. But the mechanisms of their mutagenic actions are not known. The mechanism of action of hydroxylamine (NH_2OH) is known to some extent. *Hydroxylamine induces G·C → A·T substitution most likely by inducing hydroxylation of the amino group of cytosine. This produces hydroxylaminocytosine, which base-pairs with adenine in place guanine.*

Recently, polydeoxyribonucleotides, both naturally occurring transposable elements and synthetic oligonucleotides, are becoming increasingly important for inducing mutations of specific types, usually in prokaryotes and in isolated DNA. Thus DNA is fast achieving the status of a chemical mutagen.

22.8. DETECTION OF MUTATIONS

Scoring of some types of mutations in certain organisms is relatively easy. For example, mutations for antibiotic resistance in bacteria are simply detected by plating the bacterial cells on a medium containing a lethal concentration of the concerned antibiotic (***selective medium***). The colonies that develop on such a medium will be produced by cells resistant to the antibiotic.

But detection of morphological mutations in eukaryotes requires examination of each individual of the population for the mutant phenotype. This is not only tedious requiring much time, but is also a source of errors. Therefore, elaborate procedures for mutation detection have been developed in some eukaryotes, *e.g., Drosophila,* maize, etc. These procedures employ specific ***marker genes*** (genes producing specific, distinct and easily detectable phenotypes) to facilitate the identification of chromosomes contributed to the progeny by the treated or irradiated individuals. In *Drosophila,* several genetic stocks have been constructed for the detection of lethal and visible mutations in the X-chromosome and in autosomes. *The two genetic stocks most commonly used for mutation detection in X-chromosome are,* (1) *ClB* and (2) *attached-X stocks.*

Mutation Detection

A . Drosophila

- *ClB stock*: Sex–linked recessive lethals.
- *Attached-X technique*: Sex-linked visible mutations.

B. Plants

- Homozygous recessive tester stocks.
- Any strain trated; mutations scored in M_2. (Both methods for *macromutations*)
- Any strain treated; M_1 selfed; normal plants selfed in M_2; M_3 progenies studied (*micromutations*).

22.8.1. *ClB* Stock of *Drosophila*

This technique was developed by Muller. He used it for the unequivocal demonstration of mutagenic action of X-rays. The *ClB* stock of *Drosophila* has a

special *X* chromosome a large part of which is inverted. This inversion acts as a crossover suppressor in the inverted region and is designated by *C*. A recessive lethal gene (*l*) and the dominant gene for *bar* eye (*B*) are located within the inverted segment. As a result, the *l* and *B* are inherited together in the same chromosome. Consequently, all *ClB* females (identified by bar eye shape) are heterozygous for this chromosome. Females homozygous for the *ClB* chromosome will die as they will be homozygous for the lethal gene (*l*). But males having the *ClB* chromosome do not survive. (Why ?)

The ClB technique is used for the detection of sex-linked recessive lethal mutations induced in Drosophila *males.* Mutagen-treated males are crossed with *ClB* females. In F_1, half of the females will have the *ClB* chromosome and bar eye shape.

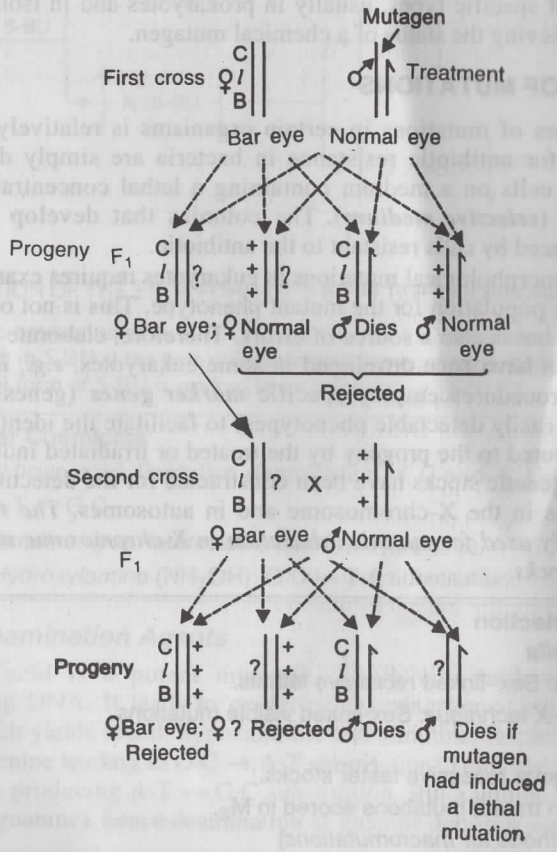


Fig. 22.5. The *ClB* method for detection of sex-linked recessive lethals in *Drosophila*. If a recessive lethal mutation is induced in an *X* chromosome of the mutagen-treated male (used in the first cross), no male progeny will be recovered from the F_1 ClB female carrying such an *X* chromosome.

The remaining half of the females will not have the *ClB* chromosome, and normal eye shape; they are rejected. All the surviving F_1 males will have the normal chromosome from the *ClB* females, while those receiving the *ClB* chromosome will die. Each F_1 *ClB* female is mated to a normal male. Progeny from each such mating are kept in separate culture bottles. Each F_1 *ClB* female will have one *ClB* chromosome; its other *X* chromosome will be from the mutagen-treated male parent. Half of the male progeny from the F_1 *ClB* female × normal male matings will receive the *ClB* chromosome, and will die. The remaining half of the males will get their *X* chromosome from their mutagen-treated grandfather. If a lethal mutation was induced in this *X* chromosome by the mutagen treatment, these males will also die (Fig. 22.5).

Thus *if the normal X chromosome of an* F_1 ClB *female had a lethal mutation, there would be no male in its progeny. The detection of sex-linked recessive lethals in this method is based only on the presence or absence of male progeny in the crosses between the* F_1 ClB *females and normal males.* The frequency of recessive lethal mutations in the *X* chromosome is obtained as the ratio between the number of F_1 *ClB* females producing no male progeny and the total number of the females tested.

$$\text{The frequency (\%) of recessive lethal mutations in the X chromosome of the mutagen-treated males} = \frac{\text{Number of } F_1 \text{ } ClB \text{ females producing no males in their progeny}}{\text{Total number of } F_1 \text{ } ClB \text{ females tested}} \times 100$$

This technique is simple, rapid and there is little chance of an error in scoring. However, it is suitable for the scoring of sex-linked recessive lethals only.

22.8.2. Attached-*X* Chromosome Technique

*This technique is based on attached-*X (X-XY) *females. It is designed to study visible sex-linked mutations in* Drosophila. Mutagen-treated males are mated with attached-*X* females. The *X-XX* (superfemale) and *YY* progeny produced from such crosses do not survive. Only the *X-XY* (female) and *XY* (male) porgeny are recovered. All the male (*XY*) progeny receive their *Y* chromosome from the attached-X female parent, while their *X* chromosome is contributed by the mutagen-treated male parent (Fig. 22.6). If a visible mutation was induced in the *X* chromosome of the sperm (produced by the mutagen-treated male), it will be expressed in the male progeny.

Therefore, *all the male progeny obtained from the cross are scored for visible mutations.* The frequency of a visible mutation is expressed as the ratio between the number of males showing a mutation and the total number of males in the progeny.

$$\text{The frequency (\%) of a visible mutation in the } X \text{ chromosome of the mutagen-treated males} = \frac{\text{Number of progeny males showing the mutation}}{\text{Total number of males in the progeny}} \times 100$$

Similarly, suitable genetic stocks have been constructed for the detection of mutations in autosomes of *Drosophila* as well. These genetic stocks use balanced lethal systems and linked dominant markers in order to identify the chromosomes derived from the mutagen-treated flies.

22.8.3. Detection of Mutations in Plants

Techniques for detection of mutations are relatively poorly developed for plant species. The following three approaches are generally adopted for this purpose.

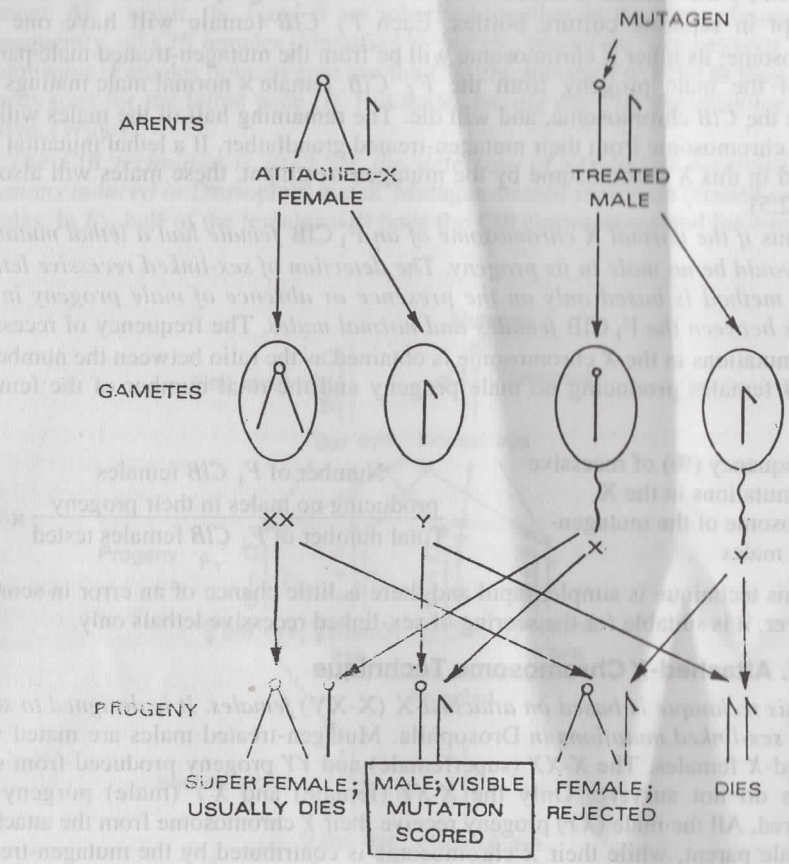


Fig. 22.6. The attached-*X* (*X-X*) method for detection of visible mutations in the *X* chromosome of *Drosophila.* All surviving males from the cross are scored for the presence of a visible mutation and the frequency of mutations is estimated on the basis of the number of surviving males.

1. In some species, *e.g.,* maize, strains homozygous for several recessive genes as well as those for dominant alleles of these genes are available. In such a case, seeds or plants of a strain homozygous for several dominant genes are treated with a mutagen. These plants (M_1, the generation treated with mutagen) are crossed with the strain having the recessive alleles of the same genes in homozygous state (***tester strain***). The treated plants are generally used as female parent due to partial male sterility in mutagen-treated plants. In the progeny of this cross, plants exhibiting recessive forms of the concerned characters are counted. The frequency of mutation for a gene is estimated as follows.

$$\text{Mutation frequency (\%) for a gene} = \frac{\text{Number of plants having the recessive form of the trait governed by the gene}}{\text{Total number of plants in the progeny}} \times 100$$

The plants showing the recessive form of the trait will receive one recessive allele from the tester parent having the recessive traits. But the other recessive allele would have been produced due to mutation in the mutagen-treated parent as this parent had the dominant allele of the gene.

2. In most plant species, however, such strains are not readily available. In such species, seeds of a variety or strain are treated with a mutagen and grown to obtain the M_1 generation. The M_1 plants are selfed to avoid outcrossing due to partial male sterility in M_1 plants. The seeds thus obtained represent the M_2 generation. M_2 plants are grown and plants having mutant features are scored. The frequency of a given mutation is estimated as per cent ratio between the number of plants exhibiting a mutation in M_2 and the total number of plants in M_2.
3. The above procedures are applicable when mutations are to be detected in qualitative traits (***macromutations***). In crop improvement, mutations in quantitative traits (***micromutations***) are equally, often more, important. For detection of micromutations, M_2 *generation is grown as described above. All* M_2 *plants showing a visible macromutation and/or partial pollen sterility are rejected. Seeds from all normal-looking fully fertile* M_2 *plants are harvested separately. They are planted in individual plant progeny rows in* M_3 *(preferably)in a replicated trial*). Data are recorded on quantitative traits. Means and variances of M_3 progenies are compared with those of the parental variety to ascertain if an M_3 progeny deviated significantly from the parental variety for a quantitative trait. A significant deviation from the parent variety will indicate the occurrence of micromutations for the concerned trait.

Molecular Basis of Mutation

- Change in base sequence of gene due to:
  - (*i*) Base substitution
    - Transition ($A \leftrightarrow G$, $C \leftrightarrow T$)
    - Transversion ($Pu \leftrightarrow Py$)

    Change in one codon: missence or nonsense mutation.
  - (*ii*) Base addition/deletion: if added/deleted base number not divisible by 3, FRAME-SHIFT MUTATION results.

22.9. MOLECULAR BASIS OF MUTATION

Any character of an organism is produced due to the proteins produced by the genes governing that trait. The properties of a protein depend on the type and the sequence of amino acids present in that protein molecule. This, in turn, is determined by the base sequence of the gene directing the production of that protein. An amino acid is coded by a set of three bases (***codon***). The base sequence of a codon determines the amino acid encoded by it.

This clearly shows that *mutation in any character is generally due to a change in the amino acid sequence of the protein determining the expression of that trait. The change in amino acid sequence of this protein would obviously be due to a change in the base sequence of the DNA molecule coding for the protein.*

The changes in the base sequence of DNA molecules may occur in one of the following two ways: (1) base substitution, and (2) base addition or deletion. In prokaryotes, both the types of changes occur spontaneously. But a surprisingly large proportion of spontaneous mutations is produced by addition or deletion of single base-pairs.

22.9.1. Base Substitution

When one base in a DNA molecule is replaced by another base, it is called ***base substitution.*** Base substitutions are of two types: (1) transition and (2) transversion. When a purine is replaced by another purine (*e.g.*, adenine is replaced by guanine or *vice-versa*) or a pyrimidine is replaced by another pyrimidine (*e.g.*, thymine is replaced by cytosine and *vice-versa*), it is known as ***transition*** (Fig. 22.7). Clearly, transition for one base affects the base sequence of only one codon. As a result, only one amino acid is altered in the concerned protein.

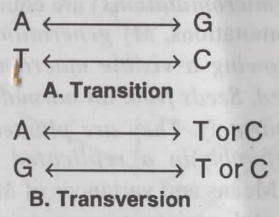


Fig. 22.7. The types of base substitutions involved in (A) transition and (B) transversion

In a ***transversion***, a pyrimidine (thymine, cytosine) is replaced by a purine (adenine, guanine) and *vice-versa*. Thus, there are eight different types of possible transversions (Fig. 22.7; A to T, A to C, G to C, G to T and *vice-versa*). Like transition, one transversion affects only one codon. Thus it replaces only one amino acid in the concerned protein. consequently, *transitions and transversions are relatively less deleterious. But sometimes, base substitutions may generate nonsense codons. Nonsense mutations are generally more deleterious than other base substitution mutations.*

22.9.2. Base Addition and Deletion

Insertion of one or more bases in a DNA molecule is called ***base addition.*** But a loss of one or more bases is known as ***base deletion.*** Both addition and deletion lead to mutations. If the number of bases added or lost is a multiple of 3, one to several amino acids would be either added to or deleted from the concerned protein. Such a change may or may not have a profound effect on the activity of the polypeptide. *But if the number of bases lost or inserted is not a multiple of three, the base sequence of all the codons beyond the point of insertion or deletion is altered* (Fig. 22.8). Consequently, all the codons beyond this point will now code for a different amino acid than before. Clearly, the reading-frame of codons is shifted in such mutations. Therefore, they are called ***frame-shift mutations.*** *Generally, proteins*

produced by frame-shift mutations are nonfunctional. As a result, such mutations are usually much more deleterious than those produced by base substitutions.

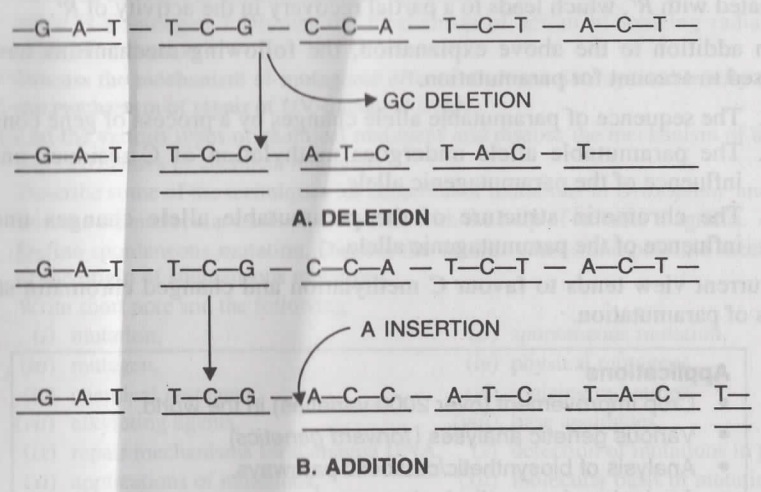


Fig. 22.8. Production of frame-shift mutations due to addition or deletion of bases in a number other than three or multiple of three. Bases of a codon are indicated by a single line drawn under them, while *the altered condons, due to a shift in the reading frame, are indicated by double lines under them. It may be noted that all the codons subsequent to (downstream of) the point of insertion or deletion are altered.*

Paramutation
- Induced by a paramutagenic allele of the gene.
- Partial loss in function of paramutated allele of the gene.
- Spontaneous recovery on separation from the paramutagenic allele.
- Due to heterochromatinization and/or DNA methylation.

22.9.3. Paramutation

Paramutation refers to a change in the function (but not the base sequence) of a structural gene (***paramutable gene***) that may last for several generations. This change is induced by the presence of the paramutable gene in the same nucleus with another allele of this gene, the ***paramutagenic gene.*** Paramutation is described in some detail in Chapter 32 (Section 32.3).

It was postulated that R^r locus has a heterochromatic segment. This segment was named as ***metamere.*** It was suggested that the metamere is associated with the structural gene *r* for anthocyanin production. The metamere acts as a suppressor of the structural gene. The degree of suppression is proportional to the number of copies of the metamere present close to the R^r allele of this gene. Allele R^r has at least one metamere, which is assumed to undergo amplification in the presence of the paramutagenic allele R^{st}. This leads to an increase in the number of metameres associated with the allele R^r. As a result, the *paramutated R^r allele is produced; this*

state has a reduced capacity for anthocyanin production. When the paramutated R^r allele is kept in the homozygous state or with the *r* allele, an under-replication of the metameres is supposed to take place. This reduces the number of metameres associated with R^r, which leads to a partial recovery in the activity of R^r.

In addition to the above explanation, the following mechanisms have been proposed to account for paramutation.

1. The sequence of paramutable allele changes by a process of gene conversion.
2. The paramutable allele undergoes methylation of C residues under the influence of the paramutagenic allele.
3. The chromatin structure of the paramutable allele changes under the influence of the paramutagenic allele.

The current view tends to favour C methylation and changed chromatin structure molels of paramutation.

Applications

- Crop improvement (over 2000 varieties) in the world.
- Various genetic analyses (*forward genetics*)
- Analysis of biosynthetic/catabolic pathways.
- Analysis of specified changes in genes on phenotypes (*reverse genetics*).

22.10. APPLICATION OF MUTATIONS

Mutations have a variety of applications, which are both basic and applied. Some of these applications are given below.

1. Some mutations have beneficial effects and are useful in ***crop improvement.*** Mutations in both qualitative and quantitative traits have been exploited for the development of over 2,000 varieties of the different crops.
2. On the fundamental side, mutation in a gene *allows that gene to be identified and studies.* For example, the *white* eye locus was identified only after the white-eyed mutant of *Drosophila* was discovered.
3. Mutations have permitted the *fine structure analysis of genes,* the determination of relationship between genes and proteins, as well as some features of the genetic code.
4. Mutations, particularly biochemical mutations, permit the *dissection of biosynthetic and biodegradation pathways* into individual reaction steps. Such an analysis would be virtually impossible in the absence of suitable mutant stocks.
5. *Finally,* induced mutations are being used for analysing the effects of various types of known alterations in DNA on the expression of the concerned characters. This is precisely the opposite of the ***classical genetics*** (also called ***forward genetics***), which ultilizes alternations in phenotypes for deducing the corresponding changes at the level of DNA. Therefore, such an analysis is often called ***reverse genetics.***

QUESTIONS

1. Define mutation. Discuss the various characteristics of mutations.
2. Describe the various classifications of mutations.
3. What is a mutagen ? Discuss the mechanism of action of ionizing radiations in mutation induction.
4. Discuss the mechanism of mutagenic effects of ultraviolet rays and briefly describe the mechanism of repair of UV-damaged DNA.
5. List the various types of chemical mutagens and discuss the mechanism of action of any two classes of chemical mutagens.
6. Describe some of the techniques for detection of mutations in *Drosophila* and plants.
7. Discuss the molecular basis of mutations with the help of suitable diagrams.
8. Define spontaneous mutation. Discuss the various causes and possible mechanisms of the origin of spontaneous mutations.
9. Write short note son the following:
   (*i*) mutation, (*ii*) spontaneous mutation,
   (*iii*) mutagen, (*iv*) physical mutagens,
   (*iv*) chemical mutagens, (*vi*) ionising radiations,
   (*vii*) alkylating agents, (*viii*) base analogues,
   (*ix*) repair mechanisms for damaged DNA, (*x*) detection of mutations in plants,
   (*xi*) applications of mutations, (*xii*) molecular basis of mutations.

23

The Genetic Material

23.1. INTRODUCTION

The chemical of which genes are composed must possess the following properties. (1) The genetic material must be replicated with high fidelity. (2) The genetic material must be able to express itself. (3) Therefore, the structure of genetic material must be able to store the highly variable information necessary for gene function. In addition, (4) the replication and/or the distribution of genetic material must allow errors in a low frequency for the origin of new genetic variation through mutation so that evolution may take place.

Genetic Material

- Material of which genes are made.
- Properties:
  1. High fidelity replication.
  2. Ability to express itself.
  3. Ability of store information.
  4. Must provide for origin of genetic variation.

23.2. IDENTIFICATION OF THE GENETIC MATERIAL

The process of identification of genetic material began in 1928 with the experiments of Griffith and concluded in 1952 with the studies of Hershey and Chase. Another ingenious experiment by Frankel-Conrat and Singer in 1957 established that in some viruses RNA functions as the genetic material.

But nuclic acids were discovered much earlier in 1871 by Meischer who called then *nuclein*. There are two types of nucleic acids, *viz.*, deoxyribose nucleic acid (DNA) and ribose nucleic acid (RNA). In eukaryotes, chromosomes contain genes, and they are make up of chromatin, *i.e.*, DNA + proteins. Obviously, either DNA or protein would be the genetic material. There was a prolonged controversy before DNA was unequivocally accepted as the genetic material.

23.2.1. Experiments of Griffith

Griffith discovered the phenomenon of transformation through his studies on *Diplococcus pneumoniae* (earlier known as *Pneumococcus*), which causes pneumonia in most of the mammals. Different strains of *Diplococcus* form one of the following two types of colonies: (1) smooth and (2) rough. The cells of strains forming smooth colonies are able to produce pneumonia and are called ***virulent***. But strains producing rough colonies are termed as ***avirulent*** since they cannot produce pneumonia.

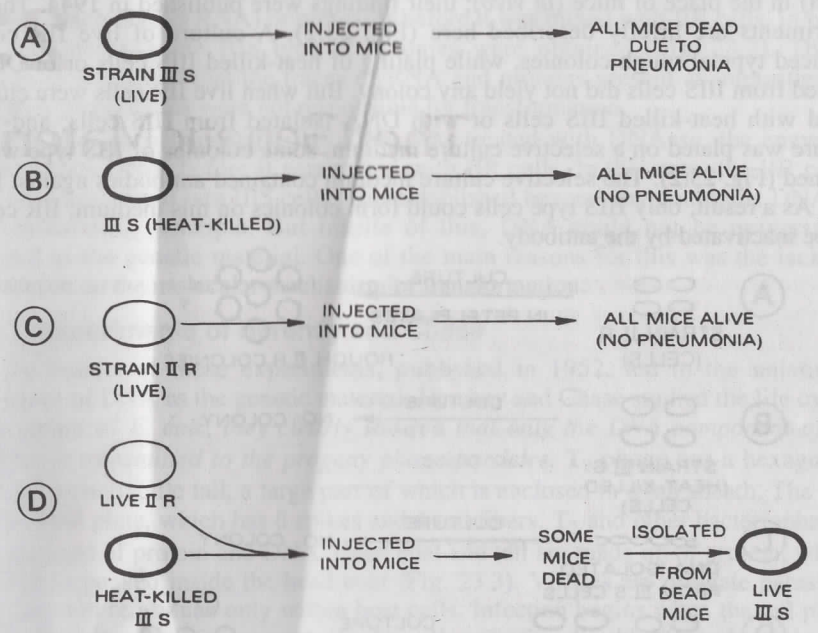


Fig. 23.1. A simplified representation of the experiments of Griffith with *Diplococcus pneumoniae.*

The experiments of Griffith (1928) are briefly described here. When live cells of the avirulent strain IIR (R is for rough colonies) were injected into mice, they did not suffer from pneumonia and all the mice survived. On the other hand, when mice were injected with live IIIS (S is for smooth colonies; virulent) cells, all the mice died due to pneumonia (Fig. 23.1A, B). Further, mice injected with heat-killed cells of the virulent strain IIIS did not develop pneumonia (Fig. 23.1*c*). However, when mice were injected with a mixture of heat-killed IIIS cells and live IIR cells, some of them died due to pneumonia. *Diplococcus* cells isolated from the dead mice were of the type IIIS (Fig. 23.1D). Since all the cells of the heat-killed IIIS culture were dead (Fig. 23.1C), *it was postulated that some of the cells of IIR changed into the IIIS type due to the influence of dead IIIS cells present in the mixture. This phenomenon was called* ***transformation***. The component of IIIS cells, which induced the conversion of IIR cells into IIIS cells was named as the ***transforming principle.***

The experiments of Griffith demonstrated transformation, but they did not hint at the identity of the transforming principle. It was later shown by Avery and coworkers that DNA is the transforming principle. ***Transformation*** may be defined, on the basis of our current knowledge, as a hereditary change in the traits of a cell/organism brought about by an integration through recombination of a segment of a DNA molecule into the chromosome of that cell organism.

23.2.2. Experiments of Avery, MacLeod and McCarty

Avery and associates carried out the experiments of Griffith *in vitro* (in a glass vessel) in the place of mice (*in vivo*); their findings were published in 1944. These experiments are briefly described here (Fig. 23.2). A culture of live IIR cells produced typical rough colonies, while plating of heat-killed IIIS cells or of DNA isolated from IIIS cells did not yield any colony. But when live IIR cells were either mixed with heat-killed IIIS cells or with DNA isolated from IIIS cells, and the mixture was plated on a selective culture medium, some colonies of IIIS type were obtained (Fig. 23.2). The selective culture medium contained antibodies against IIR cells. As a result, only IIIS type cells could form colonies on this medium; IIR cells will be inactivated by the antibody.

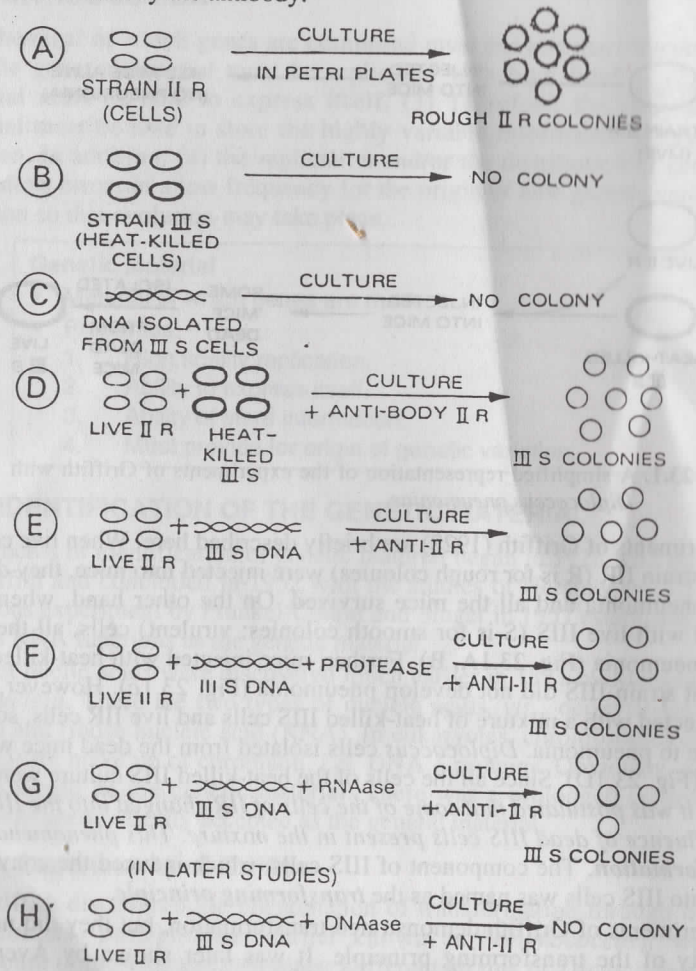


Fig. 23.2. A generalised presentation of the findings of Avery, McLeod and MacCarty with *Diplococcus*. They repeated the experiment of Griffith with certain refinements; they cultured the cells in petriplates, and used anti-IIR to precipitate IIR cells and, thereby, stop their multiplication. DNAase was used in a later study.

These findings clearly show that ***DNA is the transforming principle.*** In order to establish beyond any doubt that DNA is the transforming principle, Avery, MacLeod and McCarty, in two separate experiments, treated IIIS DNA with RNAse (the enzyme, which digests RNA) or proteases before it was mixed with IIR cells. In both the experiments, some IIIS type of colonies were obtained (Fig. 23.2). These findings clearly show that the traces of RNA and proteins present as impurities in IIIS DNA preparations were not responsible for transformation.

In a later experiment, the IIIS DNA was treated with DNAase, the enzyme, which digests DNA, before it was mixed with IIR cells. In this case, the IIR cells were not transformed into IIIS cells. This established beyond any doubt that DNA is the transforming principle. But inspite of this, DNA could not be universally accepted as the genetic material. One of the main reasons for this was the lack of information on the molecular mechanism of transformation.

23.2.3. Experiments of Hershey and Chase

The results of these experiments, published in 1952, led to the universal acceptance of DNA as the genetic material. Hershey and Chase studied the life cycle of T_2 phage of *E. coli*; *they clearly showed that only the DNA component of* T_2 *particles is transmitted to the progeny phage particles.* T_2 phage has a hexagonal head and a contractile tail, a large part of which is enclosed in a tail sheath. The tail ends in a tail plate, which has 6 spikes and 6 tail fibers. T_2 and other bacteriophages are composed of protein and DNA. Head coat and tail are made up of protein, while the DNA is packed inside the head coat (Fig. 23.3). Viruses are obligate parasites since they can reproduce only within host cells. Infection begins when the tail plate of a T_2 particle comes in contact with the cell wall of an *E. coli* cell and attaches to it. Now the tail sheath contracts pushing a part of the tail core through the bacterial cell wall.

Identification of Genetic Material

- Meischer (1871) discovered nucleic acids.
- *Griffith* (1928): Dead virulent *D. pneumoniae* change some a virulent cells into virulent form (transformation).
- Avery, MacLeod and MC Carty (1944): DNA is the transforming principle; DNA isolated from virulent cells caused transformation.
- Hershey and Chase (1952): DNA (not protein is transmitted) to progeny phages; as a result, DNA universally accepted as the genetic meterial.
- Frankel-Conrat and Singer (1957): RNA is the genetic meterial in TMV.

Hershey and Chase labelled T_2 DNA and protein in two separate experiments. DNA contains phosphorus (P) but no sulphur (S), while proteins contain S but ordinarily they have no P. Therefore, they labelled T_2 DNA with ^{32}P, while proteins were labelled with ^{35}S. In one experiment, they mixed ^{32}P-labelled T_2 particles with *E. coli* cells. Infection was allowed for 10 min. The cells were then agitated in a blender to separate the empty phage particles (called *ghosts*) remaining outside the bacterial cells after infection. The bacterial cells were pelletted out through centrifugation. Most of the ^{32}P label was present in the infected *E. coli* cells, while

the ghosts and the phage particles contained only a small fraction of the total label. The progeny phage particles obtained after lysis of these *E. coli* cells also contained the ^{32}P label (Fig. 23.3). This clearly shows that the *^{32}P present in phage DNA is injected into E.* coli *and is finally transmitted to progeny phages, and that DNA is transmitted from one generation to the next.*

In the other experiment, they used ^{35}S-labelled T_2 particles for infection of *E. coli* cells. The experimental procedure was the same as that for the ^{32}P-labelled phage particles. In this case, almost all the ^{35}S label was present in the ghosts and the active phage particles, and only a very small fraction of it was found in the infected *E. coli* cells. Further, the amount of ^{35}S present in the progeny phage particles obtained through lysis of the infected *E. coli* cells was almost negligible. *Clearly, almost all of the proteins are left outside the* E.coli *cells during infection, and that proteins are not transmitted from one phage generation to the next.*

Thus the studies of Hershey and Chase clearly showed that *DNA of T_2 is transmitted from one generation to the next, while protein is not transmitted. Therefore, at least in T_2 phage, DNA is the genetic material.* The role of DNA in these studies (transmission from one generation to the next) was clear and easily understandable. Therefore, these findings led to the universal acceptance of DNA as the genetic material; *DNA is the genetic material in all eukaryotes and most prokaryotes* (Table 23.1).

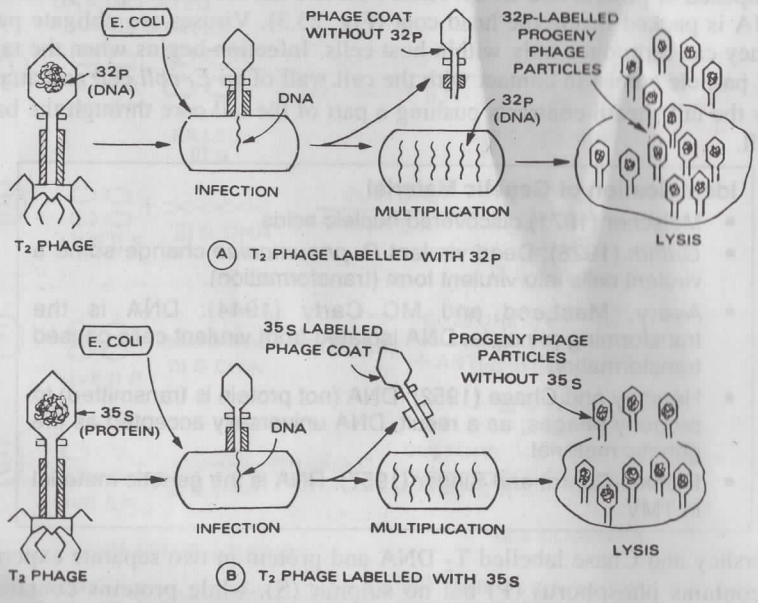


Fig. 23.3. A simplified diagrammatic representation of the experiments of Hershey and Chase with T_2 phage of *E. coli*. (*a*) When T_2 particles were labelled with ^{32}P, most of the label was recovered in the progeny phage particles, while (*b*) when they were labelled with ^{35}S, only a small fraction of the label was found in the progeny phage particles.

23.2.4. RNA as Genetic Material

In several viruses, *e.g.*, TMV (tobacco mosaic virus), DNA is absent. These viruses are composed of RNA and protein. TMV particles are like hollow cylinders. Their RNA is coiled like a spring, while the protein molecules are arranged on the outside of the coil. Frankel-Conrat and Singer demonstrated in 1957 that RNA functions as the genetic material in TMV.

Proteins and RNA of TMV can be separated chemically; when they are remixed under appropriate conditions, they reassociate to produce active TMV particles. In one experiment, Frankel-Conrat and Singer used either RNA or the proteins isolated from TMV for infection of tobacco leaves. Mosaic symptoms developed only when RNA was used for infection (and not when the proteins were used). Clearly, *only RNA fraction of TMV is capable of producing the disease, and hence appears to be the genetic material.*

In another experiment, they produced hybrid TMV particles using proteins and RNA from two different strains of the virus. Strains A and B of TMV produce

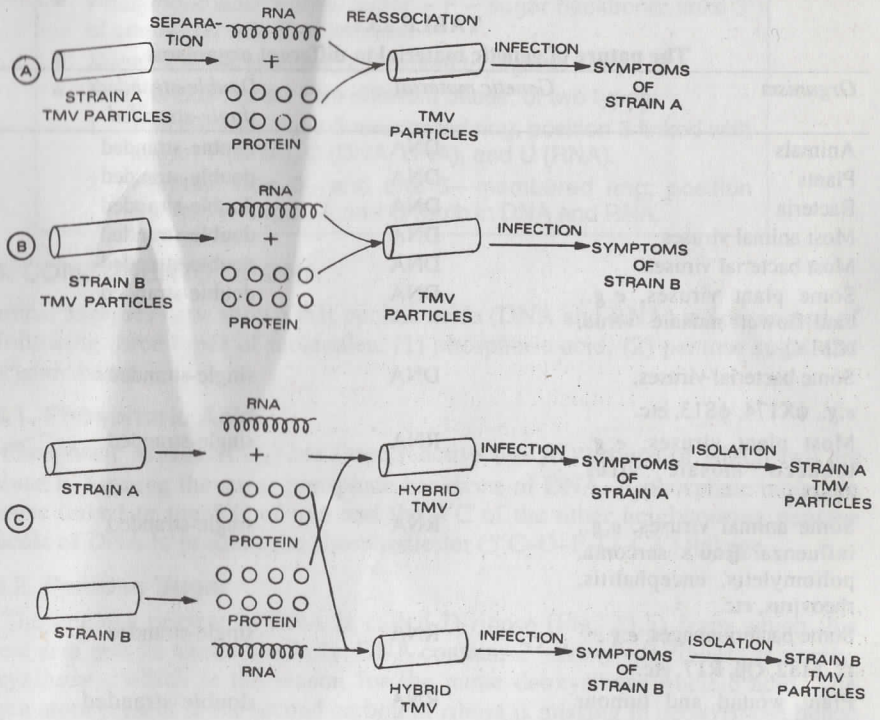


Fig. 23.4. A simplified representation of the experiments of Frankel-Conrat and Singer on tobacco mosaic virus (TMV). RNA and proteins of TMV can be chemically separated, which reassociate to form active particles of the same type as the original strain (A, B). When hybrid TMV particles were constructed by mixing RNA from one strain and proteins from another strain, the hybrid particles produced the symptoms identical with that strain, which contributed the RNA; the TMV particles isolated from the infected plants were also identical with the same strain.

distinctive symptoms on a certain tobacco variety. Frankel-Conrat and Singer constructed two types of hybrid virus particles by mixing RNA and proteins as follows: (1) RNA from strain A and proteins from strain B, and (2) proteins from strain A and RNA from strain B. When tobacco leaves were infected with hybrid TMV particles of the first type, the disease symptoms of strain A developed. Further, the virus particles isolated from the diseased leaves had proteins identical with those of strain A. Similarly, when hybrid particles of the second type were used for infection of tobacco leaves, symptoms identical with those of strain B were observed. Further, the virus particles isolated from the diseased leaves had proteins identical to those of strain B (Fig. 23.4).

It is evident from these findings that only *RNA of TMV has the capacity to produce the disease, and that the type of proteins present in the virus particles is determined by the RNA. Clearly, RNA is the genetic material in TMV.* It is now known that in many other viruses, *e.g.*, animal viruses causing poliomyelitis, influenza, encephalitis, etc., most of the plant viruses, and some small bacteriophages (such as, f2, MS2, R17, etc.) RNA functions as the genetic material (Table 23.1).

TABLE 23.1
The nature of genetic material in different organisms

Organism	*Genetic material*	*Double-stranded/ single-stranded*
Animals	DNA	double-stranded
Plants	DNA	double-stranded
Bacteria	DNA	double-stranded
Most animal viruses	DNA	double-stranded
Most bacterial viruses	DNA	double-stranded
Some plant viruses, *e.g.*, cauliflower mosaic virus, (CMV)	DNA	double-stranded
Some bacterial viruses, *e.g.*, ϕX174, ϕS13, etc.	DNA	single-stranded
Most plant viruses, *e.g.*, tobacco mosaic virus, (TMV)	RNA	single-stranded
Some animal viruses, *e.g.*, influenza, Rou's sarcoma, poliomyletis, encephalitis, rheovirus, etc.	RNA	single-stranded
Some bacteriophages, *e,g.*, f2, MS2, Qβ, R17, etc.	RNA	single-stranded
Plant wound and tumour viruses	RNA	double-stranded

23.2·5. Genetic Material in Eukaryotes

Until recently, the evidence that DNA is the genetic material in eukaryotes as well was indirect. It was based on such information that in eukaryotes genes are located in chromosomes, and that chromosomes contain DNA. The amount of DNA present in a cell is proportionate to the level of its ploidy , *e.g.*, gametes have exactly one half as much DNA as zygotes of a species. Further, the mutagenic effectiveness

of the different wavelengths of UV-rays shows the same pattern as the degree of their absorption by DNA bases. Ordinarily, DNA is the only component of the cell that is highly conserved; as a mater of fact, DNA replication normally does not lake place unless the cell prepares for division.

But it is now known that (1) RNA (all types) synthesis in eukaryotes is always associated with DNA and that (2) DNA serves as template during this process (called ***transcription***). Further, (3) the synthesis of proteins is directed by mRNA through the process of translation. Clearly, the information for synthesis of chromosomal proteins is provided by DNA. (4) Segments of eukaryotic DNA have been transferred and expressed in *E. coli* and in other eukaryotic species. These DNA segments direct the synthesis of the same mRNAs and proteins in the new species as is normally produced by them in the species from which they were taken. These findings provide direct evidence that DNA is the genetic material in eukaryotes.

Components of DNA

- *Phosphoric acid*: Forms sugar – P – sugar backbone; links 3′ of one sugar with 5′ of another.
- *Sugar*: 2′– deoxyribose; 1′ C linked to organic bases.
- *Nitrogenous bases:* Five different bases; of two types.
  1. *Pyrimidines*: Have 6-membered ring; position 3 linked with sugar; T (DNA), C (DNA, RNA), and U (RNA).
  2. *Purines*: One 6– and one 5– membered ring; position 9 linked to sugar; A and G (both in DNA and RNA.

23.3. COMPONENTS OF DNA

Chemical analyses have shown that nucleic acids (DNA and RNA) are composed of the following three types of molecules: (1) phosphoric acid, (2) pentose sugar and (3) organic bases.

23.3.1. Phosphoric Acid

Phosphoric acid (H_3PO_4) has three reactive (—OH) groups of which two are involved in forming the sugar-phosphate backbone of DNA. A phosphate molecule becomes linked to the 5′C of one and the 3′C of the other neighbouring pentose molecule of DNA to produce the phosphodiester (5′C–O–P–O–C3′) linkage.

23.3.2. Pentose Sugar

The pentose present in RNA is called D-ribose (Fig. 23.5) from which this nucleic acid gets its name. Similarly, DNA contains *2′-deoxy - D-ribose* (or simply, *deoxyribose*), which is the reason for the name deoxyribose nucleic acid. The oxygen atom present at the second carbon of ribose is missing in deoxyribose hence its name 2′-deoxyribose.

23.3.3. Organic Bases

The organic bases found in nucleic acids are also called ***nitrogenous bases***. DNA ordinarily contains four different bases called, adenine (A), guanine (G), thymine (T) and cytosine (C). RNA contains the same bases except for thymine in whose place uracil (U) is found. These five bases (A, G, T, C, U) are grouped into the following two classes: (1) pyrimidine (T, C, U) and (2) purine (A,G).

23.3.3.1. Pyrimidines

Pyrimidine bases consist of a six-membered pyrimidine ring. This is similar to the benzene ring, except that it contains nitrogen (N) in the place of carbon (C) at positions 1 and 3. The three pyrimidine bases (T, C, U) contain a keto oxygen (=O) at position 2 (Fig. 23.5). In cytosine, an amino ($-NH_2$) group is present at position 6,

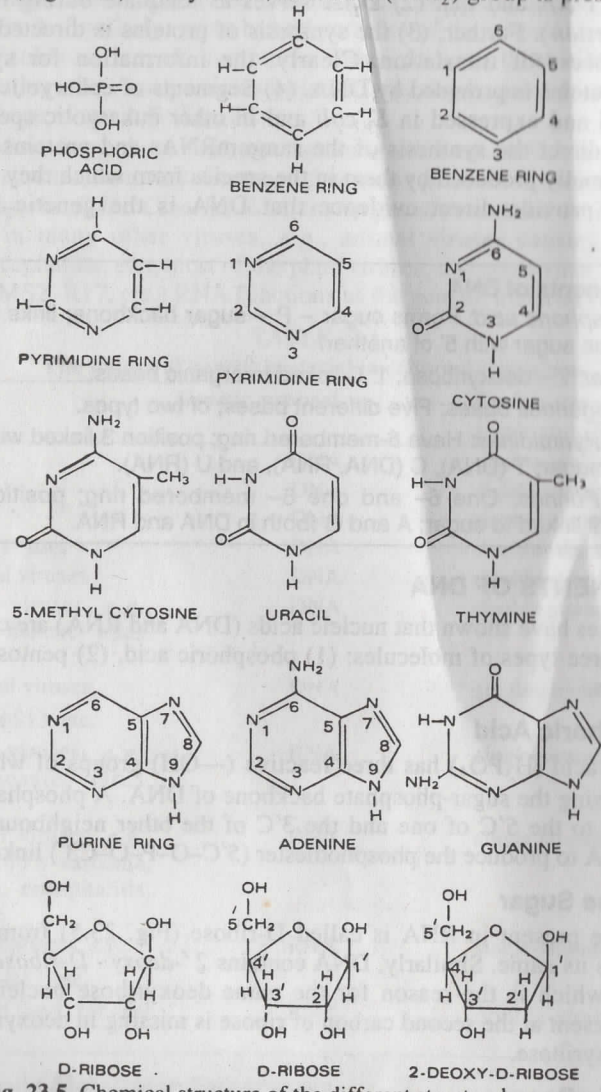


Fig. 23.5. Chemical structure of the different structural components of DNA

while in 5-methyl cytosine a methyl ($-CH_3$) group is attached to the fifth carbon. In the case of uracil, a keto (=O) group is present at the sixth carbon, while thymine is essentially 5-methyl uracil (*i.e.*, a keto oxygen at position 6, and a $-CH_3$ group at carbon 5). *All the pyrimidines, therefore, contain an –H atom at position 3, which is involved in their linkage with 1'C of pentoses.*

23.3.3.2. Purines

The two ***purines*** (A, G) consist of a purine ring, which has a six-membered and a five-membered ring. In adenine, an amino ($-NH_2$) group is present at position 6, while in guanine a keto (= O) group is found at position 6 and an $-NH_2$ group is attached with the second carbon. *The N present at the 9th position of purines participates in a covalent linkage with the 1'C of the pentose.*

23.3.4. Nucleosides

In nucleic acids, organic bases are linked with pentose molecules. In this linkage, the -H attached to the –N at position 3 of pyrimidines or that attached to the N at position 9 of purines interacts with the –OH at 1'C of pentoses, and forms a C–N covalent bond between the pentose and the organic base (Fig. 23.6). This reaction yields ***nucleosides.*** Thus nucleosides are formed by linkaged of an organic base to the 1'C of ribose/deoxyribose.

Organic base + Ribose → **Ribo**side + $1H_2O$

Organic base + Deoxyribose → Deoxyriboside + $1H_2O$

Ribosides and deoxyribosides = Nucleosides

The four common rbosides and deoxyribosides are listed in Table 23.2.

23.3.5. Nucleotides

A ***Nucleotide*** is formed when a phosphate group is attached to either the 3'C or the 5'C of the pentose molecule of a nucleoside. In the nucleotides participating in DNA replication, the phosphate groups are ordinarily attached to the 5'C. The different nucleotides are listed in Table 23.2.

Organic base + Ribose + 5′ (or 3′) phosphate → Ribotide + $2H_2O$

Organic base + Deoxyribose + 5′ (or 3′) phosphate → Deoxyribotide + $2H_2O$

Ribotides and deoxyribotides = Nucleotides

TABLE 23.2

The various bases, nucleosides and nucleotides commonly found in DNA and RNA

	DNA		*RNA*	
Base	*Deoxyriboside (base + deoxyribose)*	*Deoxyribotide (base + deoxyribose + 5′ phosphate)*	*Riboside (base + ribose)*	*Ribotide (base + ribose + 5′ phosphate)*
Adenine	Deoxyadenosine	Deoxyadenylic acid	Adenosine	5′ Adenylic acid
Guanine	Deoxyguanosine	Deoxyguanylic acid	Guanosine	5′ Guanylic acid
Cytosine	Deoxycytidine	Deoxycytidylic acid	Cytidine	5′ Cytidylic acid
Thymine	Thymidine*	Thymidylic acid	—	—
Uracil	—	—	Uridine	5′ -Uridylic acid

* The cells ordinarily contain only deoxythymidine and deoxythymidylic acid; hence, they are simply called thymidine and thymidylic acid, respectivley.

23.4. PRIMARY STRUCTURE OF DNA

A native DNA molecule is double-stranded. Each of the two strands of a DNA molecule has many deoxyribonucleotides, and is known as a ***polynucleotide***. These

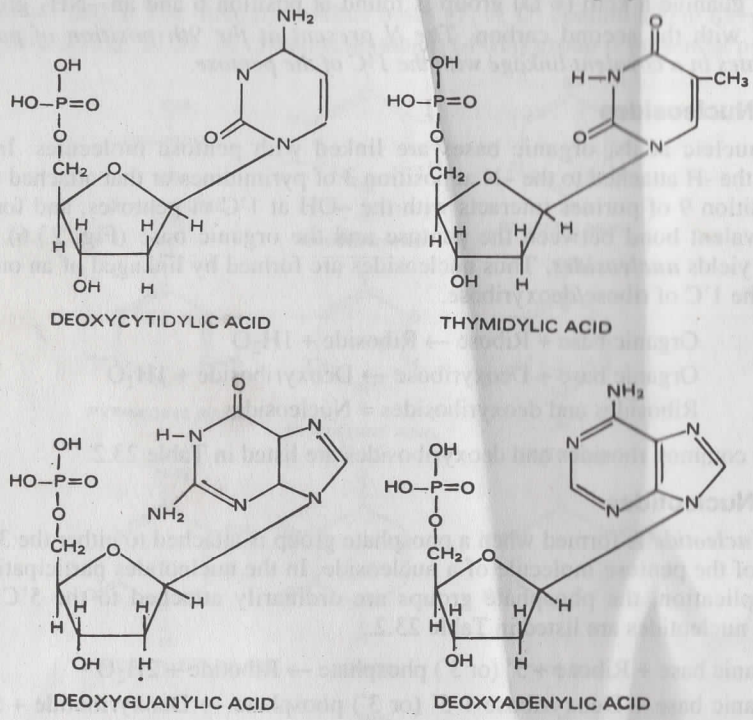


Fig. 23.6. Chemical structure of the four nucleotides commonly fond in DNA. Nucleosides have the same structure minus the phosphate residue (the – CH_2 group to which the phosphate residue is attached in the nucleotides has a –OH group in nucleosides).

nucleotides are joined with each other by phosphodiester linkages. In a ***phosphodiester linkage***, the 5′C of the pentose of one nucleotide is linked with one –O–of the phosphate, while the 3′C of pentose of the other nucleotide is linked with another –O– of the same phosphate residue. This produces a 5′C–O–P–O–C3′ linkage. In a dinucleotide, the 5′C of one nucleotide (linked with phosphate) and the 3′C of the other nucleotide have free –OH groups with which other nucleotides can join to produce a polynucleotide; this is the ***primary structure*** *of DNA* (Fig. 23.7).

Thus a polynucleotide chain consists of a backbone made-up of several alternating pentose and phosphate molecules linked with each other; this is called ***sugar-phosphate backbone***. Each pentose has any one of the four organic bases attached to its 1′C. The bases are not directly involved in the formation of sugar-phosphate backbone of a polynucleotide. At one end of the polynucleotide chain, the 5′C of pentose has a free –OH (–OH of the phosphate group attached to the 5′C), while the 3′C of the pentose at the opposite end has a free –OH; these ends are known as *5′- and 3′-ends*, respectively. *Thus the primary structure of DNA does not impose any restriction on the sequence in which the bases occur in a polynucleotide.*

23.5. THE DNA DOUBLE HELIX

The deoxyribonucleotides and their ability to from polynucleotide chains were discovered by Levene in 1931. Chemical analyses of DNA by Chargaff to and others during 1940s clearly demonstrated the following features (Table 23·3).

1. As a rule, the number of pyrimidine bases, *i.e.*, C+T, is equal to that of purine bases, *viz.*, A + G.
2. Similarly, A + C content of DNA is equivalent to that of G + T.
3. The quantity of A is always equal to that of T, while the quantity of G is equal to that of C.

TABLE 23.3
Haploid DNA contents and base compositions of DNAs from some organisms

Organism	*Haploid DNA content (pg = 10^{-12} g)*	*Purines*		*Pyrimidines*		*(A+G)/(C+T) ratio*
		Adenine (%)	*Guanine (%)*	*Cytosine (%)*	*Thymine (%)*	
		Animals				
Man	3.2	31.0	19.1	18.4	31.5	1.00
Cattle	2.8	28.7	22.2	22.0	27.2	1.03
Chicken	1.3	28.0	22.0	21.6	28.4	1.00
Frog (*Rana pipiens*)	6.5	26.3	23.5	23.8	26.4	0.99
D. melanogaster	0.18	27.3	22.5	22.5	27.6	0.99
		Plants				
Maize	7.5	25.6	24.5	24.6	25.3	1.00
Tobacco	1.2	29.7	19.8	20.0	30.4	0.98
Neurospora crassa	0.046	23.0	27.1	26.6	23.3	1.00
		Bacteria				
Diplococcus pneumoniae	0.002	30.3	21.6	18.7	29.5	1.08
E. coli	0.0047	24.6	25.5	25.6	24.3	1.00
Bacillus subtilis	0.003	28.4	21.0	21.6	29.0	0.98
		Viruses				
Bacteriophage T2	0.0002	32.6	18.2	16.6	32.6	1.03
Bacteriophage λ	0.000055	26.0	23.8	24.3*	25.8	0.99
Bacteriophage ϕX174**	0.0000026	24.7	24.1	18.5	32.7	0.95
Bacteriophage fd**	0.0000022	24.4	19.9	21.7	34.1	0.79

* Present as 5-hydroxymethyl cytosine.

** In these viruses, A is not equal to T and G is not equal to C as they have single-stranded DNA.

The data of Wilkins, Franklin, and others on X-ray crystallography of purified DNA revealed it to have the following features: (1) multistranded molecule, (2) diameter of ~ 22 Å, (3) groups spaced at 3·4 Å along its length, and (4) a repeating unit occurring at every 34 Å.

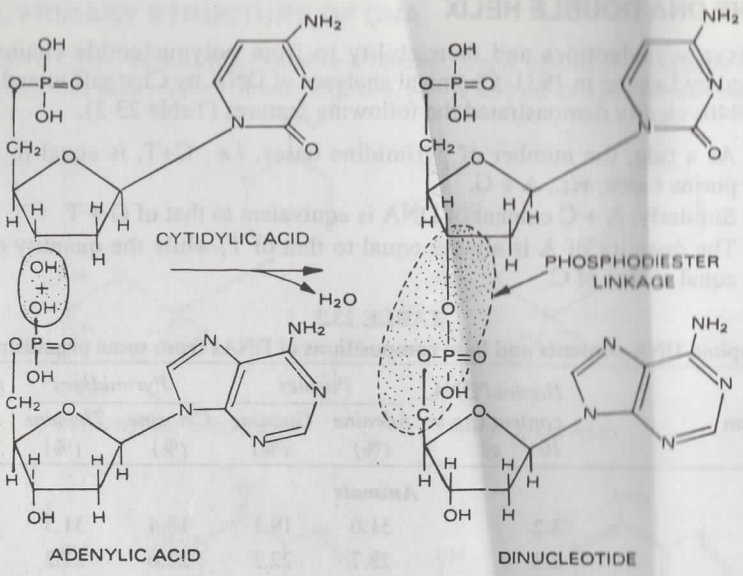


Fig. 23.7. Formation of a phosphodiester linkage between two nucleotides, cytidylic acid and adenylic acid.

DNA Double Helix: Basis

- *Levene* (1931): Discovered nucleotides.
- *Chargaff and others* (1940s): Chemical analyses of DNA revealed:
  1. A + G = C + T
  2. A + C = G + T
  3. A = T and G = C.
- *Wilkins, Franklin and others*: X-ray crystallography of DNA revealed:
  1. multistranded molecule
  2. diameter ~ 22 Å
  3. groups spaced at 3·4 Å
  4. repeating unit 34 Å.
- *Watson and Crick (1953)*: proposed the double helix model on the basis of above findings.

These findings were available by early 1950s, and were used by Watson and Crick to develop the double helix model of DNA formally proposed in 1953. This model was soon universally accepted. The main features of this model are summarised below.

1. A DNA molecule is made up of two polydeoxyribonucleotide (or simply polynucleiotide) strands or chains (Fig. 23.8).
2. The two strands of a DNA molecule are oriented antiparallel to each other, *i.e.*, the 5′-end of one strand is located with the 3′-end of the other strand at

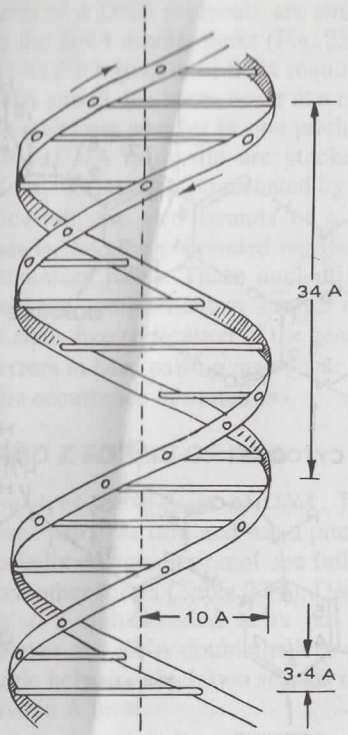


Fig. 23.8. The double helix model of DNA molecule as proposed by Watson and Crick in 1953. The cross-bars connecting the two longitudinal axes are made up of complementary bases (A : T and G : C), while the longitudinal axes are composed of alternating molecules of deoxyribose and phosphoric acid.

the same end of a DNA molecule. This antiparallel orientation of the two strands is essential for the formation of hydrogen bonds between the pairs of DNA bases (Fig. 23.9).

3. Each polynucleotide strand is composed of many deoxyribonucleotides.
4. The base sequences of the two strands of a DNA molecule show the following universal relationship. (*i*) Wherever adenine occurs in one strand, thymine is present in the corresponding position of the other strand and *vice-versa.* Similarly, (*ii*) the sites at which guanine is present in one strand are occupied by cytosine in the second strand and *vice-versa*. Therefore, the two strands of a DNA molecule are called ***complementary strands.*** Thus, if the base sequence of one strand of DNA is known, the base sequence of its complementary stand can be easily deduced.
5. The adenine present in one strand of a DNA molecule is linked by two hydrogen bonds with the thymine located opposite to it in the second strand, and *vice-versa*. Similarly, G located in one strand forms three hydrogen bonds with the C present opposite to it in the second strand, and

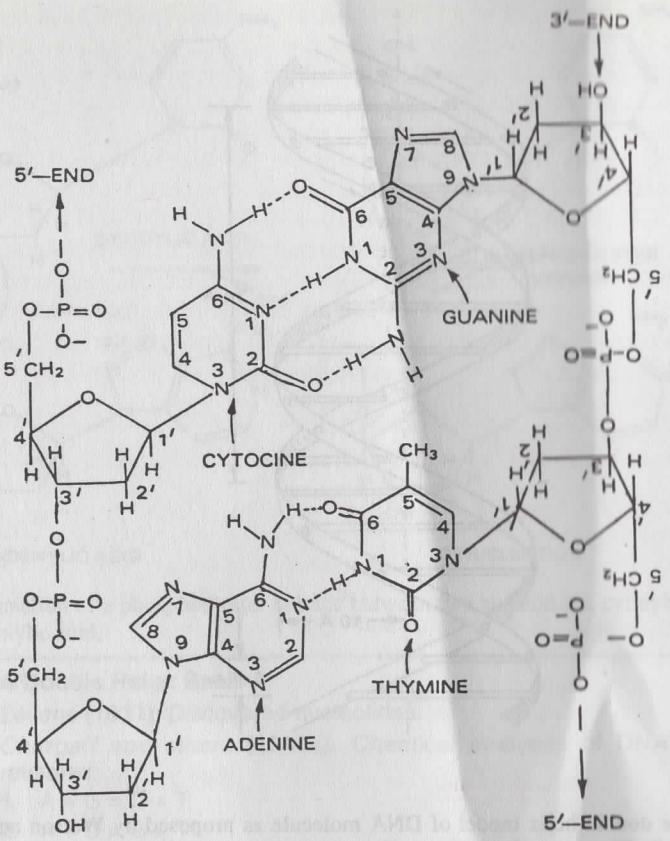


Fig. 23.9. The antiparallel orientation of the two strands of DNA molecule. AT the 3′-end of a strand the 3′ carbon of the deoxyribose is present; usually it has a free –OH group. Similarly, at the 5′-end the 5′ carbon of the pentose is present (generally, it has a phosphate residue). The complementary bases of the two strands are joined together by hydrogen bonds (H bonds), which are shown here by broken lines (...).

vice-versa. The formation of hydrogen bonds between A and T, and between G and C is known as ***complementary base pairing***. These base pairs are called ***complementary bases***. The hydrogen bonds keep the two strands of a DNA molecule together.

DNA Double Helix: Features

- A DNA molecule has two strands in antiparallel orientation, held together by H bonds between the bases.
- *A pairs with T, and G pairs with* C (complementary bases).
- A : T pair forms 2H bonds, G : C pair forms 3H bonds.
- Helix coiled in right hand direction; pitch = 34 Å; diameter 20 Å; bases occur at every 3·4 Å; 10 bases per turn.
- Replication occurs by strand separation; new strand formed by complementary base pairing.

6. The two strands of a DNA molecule are coiled together in a right-handed helix forming the ***DNA double helix*** (Fig. 23.8). The diameter of this helix is 20 Å, while its pitch (length of helix required to complete one turn) is 34 Å. In each DNA strand, the bases occur at a regular interval of 3.4 Å so that about 10 base pairs are present in one pitch of a DNA double helix. The base pairs in a DNA molecule are stacked between sugar phosphate backbones of the two strands (represented by cross-bars in Fig. 23.8).
7. During replication, the two strands of a DNA molecule uncoil. The unpaired bases in the single-stranded regions of the two strands pair with their complementary bases. These nucleotides become joined with each other and yield the complementary strands of the old ones. This provides for an almost error-free replication of the genetic material.
8. Sometimes, errors in base-pairing may occur during replication. This would account for the occurrence of mutations.

23.6. THE A, B, C AND Z FORMS OF DNA

The above description is of the ***B-from of DNA***. This form shows right-handed coiling, contains 10 base pairs per turn and has a pitch of 34Å. ***Pitch*** is the vertical distance travelled by a helix on completion of one full turn of the helix. DNA is also known to occur in three other forms (Table 23·4). Dehydrated DNA occurs in the ***A from***; this form is also a right-handed helix but has 11 base pairs per turn. DNA·RNA heteroduplexes and RNA double helices occur *in vivo* in the ***A from***. A ***heteroduplex*** is a double helix in which two strands differ from each other. DNA is unlikely to occur *in vivo* in A form.

TABLE 23.4
Main differences between B-form and Z-form DNAs

Characteristic	*A-DNA*	*B-DNA*	*C-DNA*	*Z-DNA*
Coiling	Right-handed	right-handed	right-handed	left-handed
Pitch	Ca. 28 Å	Ca. 34Å	Ca. 31 Å	Ca. 69 Å
Base pairs per turn	11	10 (10-16.6)	9.33	12
Diameter	23 Å	19 Å	19 Å	18 Å
Vertical rise per base pair	2.56 Å	3.38 Å	3.32 Å	5.71 Å
Sugar-phosphate backbone	regular	regular	regular	zig-zag

The ***Z-from of DNA*** shows left-hand coiling, contains 12 base-pairs per the turn and has a pitch of 69 Å. In addition, the sugar-phosphate backbone follows a zig-zag path, which gives the name of this form. Z-DNA was discovered *in vitro* and is not known *in vivo*. ***C-DNA*** is a right-handed helix that has 9·33 base-pairs per turn, 19 Å diameter and ~ 31 Å pitch. It is more tightly wound than the B-form DNA.

23.7. STRUCTURE OF RNA MOLECULES

RNA, like DNA, is a polynucleotide. It is produced by phosphodiester linkages between ribonucleotides in the same manner as in the case of DNA. RNA nucleotides have ribose sugar (in place of deoxyribose in DNA), which participates in the formation of sugar-phosphate backbone of RNA. Thymine is usually absent in the RNA, and uracil is found in its place. Usually, RNA is single-stranded, but double-

stranded RNA is also found. In most organisms, RNA performs nongenetic functions, *e.g.*, RNA, rRNA, and mRNA, but in some viruses it serves as the genetic material. A comparative study of DNA and RNA is summarised in Table 23.5.

TABLE 23.5
Comparison between DNA and RNA

Characteristics	*DNA*	*RNA*
Pentose	***Deoxyribose****	***Ribose***
Base	Ordinarily, ***thymine present and uracil absent***	Ordinarily, ***uracil present and thymine absent.***
Number of strands	Generally, double-stranded	Generally, single-stranded
Function	Genetic material only	(*i*) Generally, nongenetic functions, *e.g.*, mRNA rRNA, tRNA, etc. (*ii*) In some viruses, genetic material
Origin	(*i*) Replication of preexisting DNA (*ii*) In case of infection by RNA viruses, reverse transcription of genetic RNA	(*i*) ***Genetic RNA*** either through transcription of DNA, or (*ii*) Through replication of RNA by RNA-dependent RNA ploymerase (*iii*) ***Nongenetic RNA*** from transcription of DNA.
Native form	Double-stranded DNA usually in B-form	Double-stranded RNA usually in A-form

*Main differences are depicted in ***dark letters.***

QUESTIONS

1. Define genetic material and briefly describe its properties. Describe one experiment, which clearly showed that DNA is the genetic material.
2. Describe the experiment, which demonstrated that RNA is the genetic material in TMV. List the differences between DNA and RNA.
3. Describe the double-helix model of DNA structure with the help of a suitable diagram.
4. Write short notes on the following:
   (*i*) transformation, (*ii*) organic bases,
   (*iii*) nucleosides and nucleotides, (*iv*) differences between DNA and RNA,
   (*v*) primary structure of DNA, (*vi*) complementary base paring,
   (*vii*) different forms of DNA, (*viii*) genetic material in eukaryotes.

24

Replication and Repair of Genetic Material

24.1. INTRODUCTION

The genetic material must produce a large number of copies of itself during the life cycle of an organism, that is, it must undergo ***replication.*** The process of replication must be such that it faithfully reproduces the information present in the DNA undergoing replication. At the same time, it must allow a very small degree of error for mutations to occur so that new genetic variation continues to arise. This aspect is extremely important since origin of genetic variation is a prerequisite for organic evolution to occur.

In addition, DNA like any other biochemical is liable to be damaged by various physicochemical factors of the environment. These damages must be repaired, as far as possible, without causing a change in the genetic information contained in the DNA. In this chapter, we consider in some detail the processes of DNA replication and repair.

24.2. DNA REPLICATION

Theoretically, there are the following three possible modes of DNA replication: (1) dispersive, (2) conservative and (3) semi-conservative. In ***dispersive replication***, the parent DNA molecule is broken into several pieces, and the 'new' molecules will consist of both old and newly synthesized segments. In the ***conservative scheme***, the two '*old*' strand would be present in one daughter molecule while the two newly produced strands would form the second daughter molecule. But in ***semiconservative mode of replication***, each daughter DNA molecule consists of one 'old' and one 'new' strand.

All the available evidence clearly indicates that *DNA replication is semiconservative*. In ***semiconservative DNA replication***, each newly synthesized strand of DNA remains associated with the old strand against which it was synthesized. Thus each progeny DNA molecule would consist of one '*old*' and one '*newly synthesized*' strand. The semiconservative mode of DNA replication was postulated by Watson and Crick along with their double-helix model of DNA. The DNA molecule that undergoes replication may be termed as '***parent molecule***' or '***template molecule***', while the two molecules produced by replication may be called '***progeny molecules***' or '***daughter molecules***'.

The main features of this model of DNA replication are as follows (Fig. 24.1: (1) progressive separation of the two strands of the DNA molecule undergoing replication, (2) complementary base-pairing of the bases located in the single-

stranded regions thus produced with the appropriate free deoxyribonucleotides, and (3) formation of phosphodiester linkages between the neighbouring deoxyribonucleotides that have base-paired with the single strands, thereby producing the new strand. (4) This ensures that the base sequences of the new strands are strictly complementary to those of the old strands and that (5) each DNA molecule produced by replication has one 'old' and one 'new' strand. Clearly, the old strands of a DNA molecule serve as ***templates*** or ***molds*** for the synthesis of new strands.

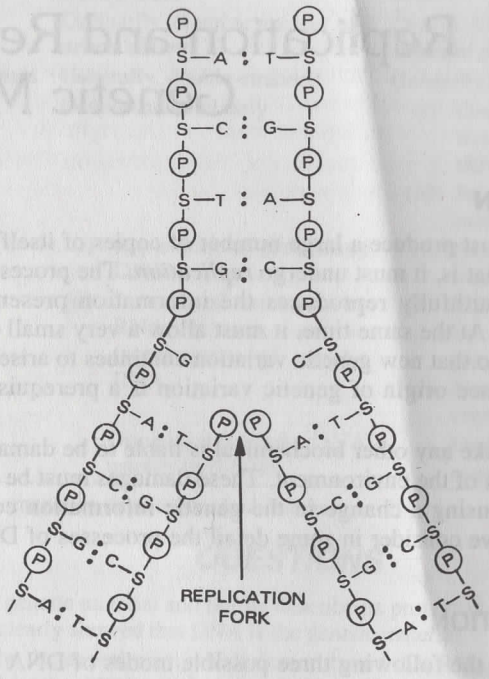


Fig. 24.1. Flawless replication of DNA through complementary pairing between DNA bases. The two DNA strands begin unwinding at one end from which the replication is initiated. This produces a Y-like configuration called replication fork. The two complementary strands go on unwinding as the replication progresses.

DNA Replication

- Semiconservative replication:
  1. The two strands of the parent DNA molecule separate.
  2. Complementary strands synthesized using 'old' strands as template.
  3. Each progeny strand has one 'old' and one 'new' strand.
- Originally suggested by Watson and Crick.
- Evidence provided by Meselson and Stahl (1958). They studied DNA replication in *E. coli* in presence of ^{15}N.
- Cairns (1963) isolated complete replicating *E. coli* chromosome, which forms a θ–shaped structure.
- Replication is bidirectional; easily shown in T_7 chromosome replication.

24.2.1. Evidence for Semiconservative Replication

The evidence for semiconservative replication of DNA was first presented by Meselson and Stahl in 1958. They grew *E. coli* on ^{15}N (a heavy isotope of ^{14}N) so that the nitrogen present in DNA bases of these cells was ^{15}N. DNA having ^{15}N has a detectable higher density than that having ^{14}N; therefore, they are called ***heavy*** and ***light DNA***, respectively. Heavy and light DNAs can be readily separated through equilibrium density gradient centrifugation; they form distinct bands in the centrifuge tube (Fig. 24.2). In density gradient centrifugation, heavy salt, *e.g.*, Cs Cl_2, is centrifuged at high speed (30,000–50,000 rpm) for 40 – 72 hr. This leads to the formation of a linear gradient of increasing density from the top to the bottom of the centrifuge tube. When molecules with small differences in density are now centrifuged in this solution, they form separate bands as shown in Fig. 24.2.

Meselson and Stahl transferred the *E. coli* cells grown on ^{15}N medium to a medium containing normal ^{14}N. They withdrew samples from these *E. coli* cells after approximately one, two and three cell generations. ***One cell generation*** represents the time during which all the cells present in a culture would undergo, on an average, one cell division. DNA from these cell samples was isolated and subjected to density gradient centrifugation. The results obtained from their study may be summarised as follows (Fig. 24.2). After one cell generation, the DNA formed a single band 'intermediate' between the 'heavy' (containing ^{15}N) and 'light' (containing ^{14}N) DNAs. The DNA obtained after two cell generations formed two bands: one of the bands was 'intermediate', while the other was 'light' in density. The same two bands were recovered in the DNA isolated after three cell generations,

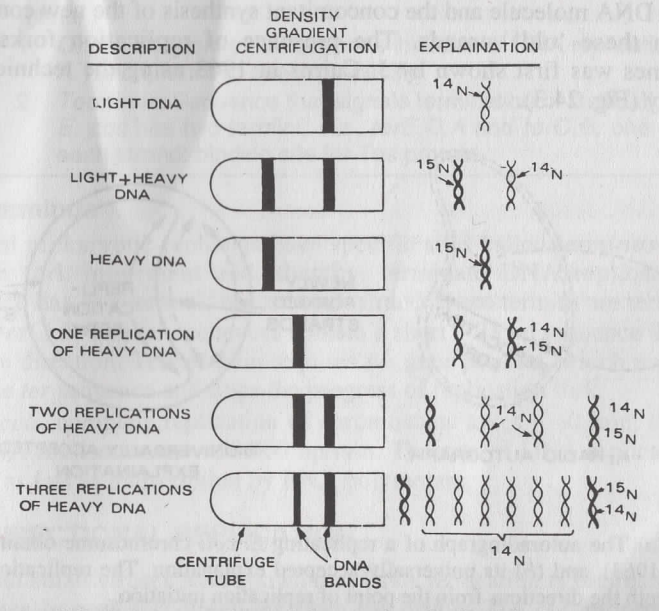


Fig. 24.2. A simplified representation of the results obtained by Meselson and Stahl by growing ^{15}N–labelled *E. coli* cells on ^{14}N (normal N) medium for different periods. The most plausible explanation for their findings is presented at the right.

although the 'intermediate' band was relatively lower in intensity than the 'light' band.

These findings can be readily explained on the basis of semiconservative replication of DNA (Fig. 24.2). The DNA from *E. coli* cells grown on ^{15}N had ^{15}N in both the strands; therefore it was heavier than the normal DNA having ^{14}N. When these *E. coli* cells were allowed one semiconservative replication of their DNA on the ^{14}N medium, each of the resultant DNA molecules would have one 'heavy' (having ^{15}N) and one 'light' (having ^{14}N) strand. Therefore, these DNA molecules would have 'intermediate' density. One more semiconservative replication of these DNA molecules in the ^{14}N medium would generate two types of DNA molecules: (1) half of the molecules would have one 'heavy' and 'one light' strand ('intermediate' density), while (2) the remaining half would have both 'light' strand ('light' density). These molecules would obviously form one 'intermediate', and one 'light' band. Following the third round of DNA replication on ^{14}N, the 'intermediate' density DNA molecules would yield half 'intermediate' and half 'light' molecules, while all the molecules obtained from replication of the 'light' DNA would be 'light'; this is the reason for the lower intensity of the 'intermediate' band after three cell generations.

24.2.2. Autoradiograph of A Replicating *E. coli* chromosome

One of the consequences of semiconservative replication is the presence of a replication fork in a DNA molecule undergoing replication (Fig. 24.3). At the replication fork, one DNA molecule appears to be branched into two molecules. Such a fork is formed due to a progressive separation of the two strands of a replicating DNA molecule and the concomitant synthesis of the new complementary strands on these 'old' strands. The presence of replication forks in *E. coli* chromosomes was first shown by J. Cairns in 1963 using the technique of autoradiography (Fig. 24.3).

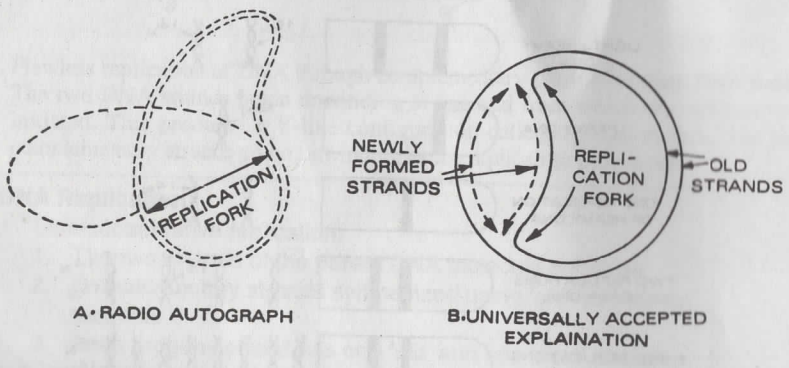


Fig. 24.3. (*a*) The autoradiograph of a replicating *E. coli* chromosome obtained by Cairns (1963), and (*b*) its universally accepted explanation. The replication proceeds in both the directions from the point of replication initiation.

E. coli chromosomes were shown to exist as θ-shaped structures during replication (fig. 24.3). It was also clear that the unwinding of the two strands of a replicating DNA molecule and the synthesis of the new complementary strands

occur almost simultaneously. The semiconservative replication of DNA begins at a definite point called 'origin' and proceeds in both the directions; thus both the Y-shaped structures in Fig. 24.3 are replication forks.

24.3. REPLICON

A ***replicon*** is that segment of DNA that is capable of DNA replication independent of other segments of DNA. Therefore, each replicon has an ***origin of replication*** (or simply, ***origin***) at which DNA replication begins, and it may have a ***terminus*** at which replication stops. The bacterial and viral chromosomes usually contain a single replicon per chromosome. But in eukaryotes, each chromosome is made up of several replicons, *e.g.*, a total of 3,500 replicons in the 4 chromosomes of *Drosophila.* The replicon size varies from 40 kb in *Drosophila* to 300 kb in *Vicia.* However, usually, only a proportion (~15%) of replicons function during replication in S phase.

24.3.1. Origin

The sequence of a replicon that supports initiation of DNA replication is called ***origin***. Origin also controls the frequency of replication of the replicon. In *E. coli*, origin is identified as the *oriC* locus, which is 245 bp long. In general, origins are A.T. rich, which may be important for easy unwinding of the two strands during replication. Origins function in a species-specific manner in that origin of one species functions well in that species or a group of related species.

Replicon
- DNA segment capable of independent replication.
- It has:
  1. *Origin*: Sequence that supports initiation of DNA replication; A.T.-rich; function species-specific.
  2. *Terminus*: Sequence that signals termination of replication; *E. coli* has two termini, *viz.*, *terE,D,A* and *terC,B*, one on each strand; binding site for Tus protein.

24.3.2. Terminus

Several prokaryotic replicons have specific sites called ***terminus***, which stop replication fork movement and, thereby, terminate DNA replication. *E. coli* chromosome has two termini, one on each strand; these termini are termed as *terE, D, A* and *terC, B*. The *ter* sequences contain a short ~ 23 bp sequence that functions in only one direction. Termination requires *tus* gene product, which recognizes and binds to the *ter* sequence and stops the progress of replication fork.

In *E. coli*, complete replication of chromosome takes ~ 40 min; this gives the rate of DNA replication as ~ 50,000 bp/min. The rate of DNA replication appears ~10-times as fast as transcription by RNA polymerase.

24.4. BIDIRECTIONAL REPLICATION

Studies with small viruses have shown that DNA replication is bidirectional, that is, it progresses in both the directions from the origin. The chromosome of *E. coli* phage T_7 replicates in a linear form, and DNA replication begins near one end of the chromosome. Replication produces the so-called eye-structure in the T_7

chromosome (Fig. 24.4). If the replication were bidirectional, the replication fork will reach one end of the chromosome much sooner than the other end. This will give rise to a Y-shaped chromosome. Both eye-shaped and Y-shaped chromosomes of T_7 chromosome have been seen under electron microscope. This confirms the DNA replication to be bidirectional. DNA replication in eukaryotes is also bidirectional.

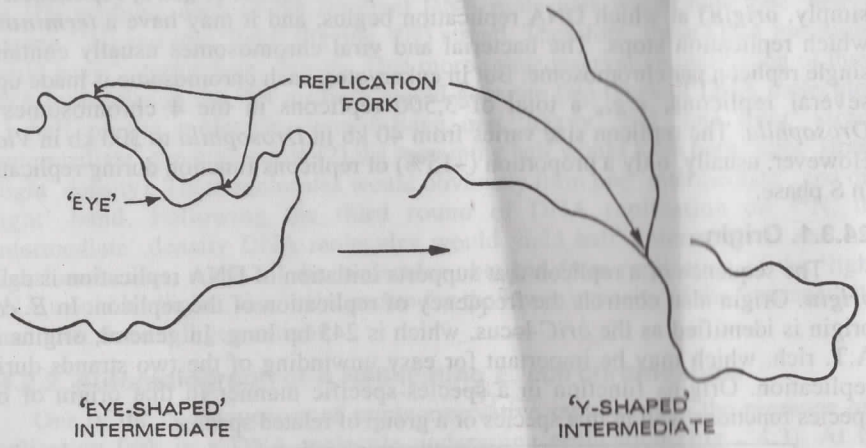


Fig. 24.4. Replicating linear chromosome of *E. coli* phage T_7 as seen under electron microscope. Replication begins near one end generating the eye-shaped intermediate. When one of the forks reaches the nearest end, a Y-shaped intermediate results.

DNA Replication: Generalized Model

- DNA undividing due to gyrase and helicase.
- Primase synthesizes RNA primer .
- DNA polymerase III elongates primer in *E. coli.*
- DNA strand oriented in 3′→ 5′ direction from origin replicates continuously (***leading strand***).
- Strand oriented in 5′→ 3′ direction replicates discontinuously (***lagging strand***); primers synthesizes primase at certain intervals.
- Replication of lagging strand produces small Okazaki fragments.
- Primer removed and gap filled by DNA polymerase I; nick sealed by ligase.

24.5. A GENERALISED MODEL FOR DNA REPLICATION

DNA replication is catalysed by the enzyme DNA polymerase, and it involves several other proteins/enzymes. The data collected from both *in vitro* and *in vivo* studies have been synthesized into a generalised model of DNA replication, a brief and simplified description of which is given below (Fig. 24.5).

1. DNA replication begin at certain unique and fixed points called '*origin*'.

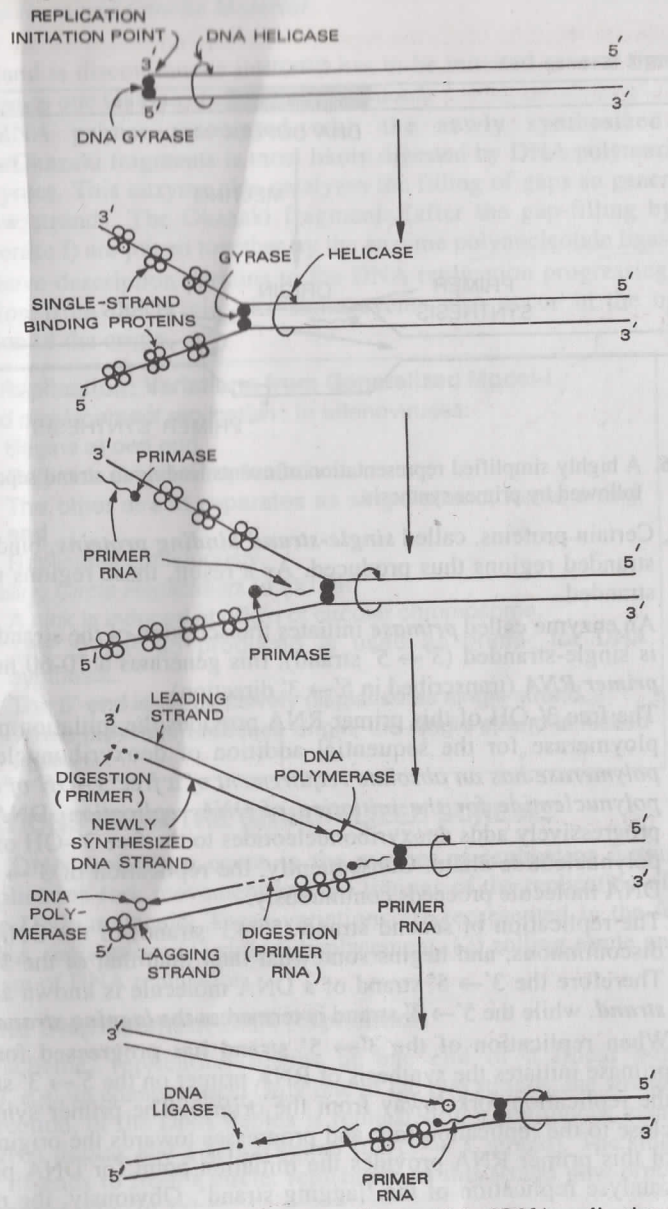


Fig. 24.5. A widely accepted generalised model of DNA replication

2. Two enzymes ***DNA gyrase*** and ***DNA helicase***, bind to the origin points and induce the unwinding of complementary strands of DNA double helix; this is known as ***DNA melting*** (Fig. 24.6).
3. Melting produces two Y-shaped forks at origin; one fork is located at each end of the origin. When replication begins, these forks become ***replication forks*** (Fig. 24.6).

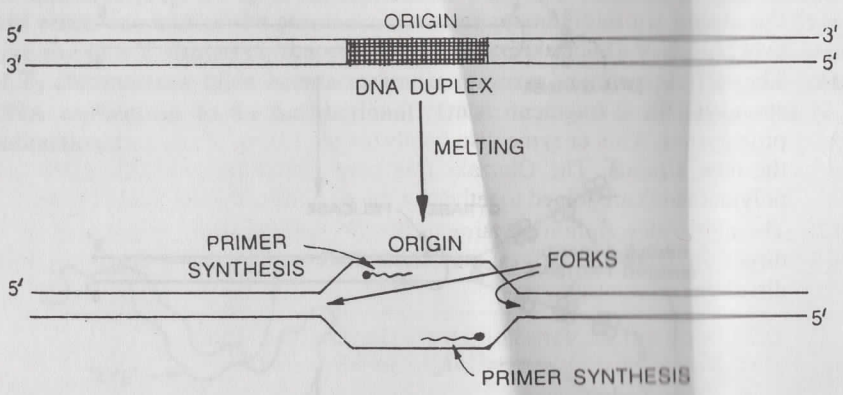


Fig. 24.6. A highly simplified representation of events leading to strand separation at origin, followed by primer synthesis.

4. Certain proteins, called ***single-strand binding proteins***, bind to the single-stranded regions thus produced. As a result, these regions remain single-stranded.
5. An enzyme called ***primase*** initiates transcription of the strand whose 3′-end is single-stranded (3′→ 5′ strand); this generates a 10-60 nucleotide long ***primer RNA*** (transcribed in 5′→ 3′ direction).
6. The free 3′-OH of this primer RNA provides the initiation point for DNA ploymerase for the sequential addition of deoxyribonucleotides. *DNA polymerase has an absolute requirement of a free 3′–OH of a pre-existing polynucleotide for the initiation of DNA replication.* DNA polymerase progressively adds deoxyribonucleotides to the free 3′–OH of this growing polynucleotide chain. Consequently, the replication of 3′→ 5′ strand of a DNA molecule proceeds continuously.
7. The replication of second strand (5′→3′ strand) of the DNA molecule is discontinuous, and begins somewhat later than that of the 3′→ 5′ strand. Therefore the 3′→ 5′ strand of a DNA molecule is known as the ***leading strand***, while the 5′→ 3′ strand is termed as the ***lagging strand***.
8. When replication of the 3′→ 5′ strand has progressed for some time, primase initiates the synthesis of RNA primer on the 5′→ 3′ strand close to the replication fork (away from the origin). The primer synthesis begins close to the replication fork and progresses towards the origin. The 3′–OH of this primer RNA provides the initiation point for DNA polymerase to catalyse replication of the 'lagging strand'. Obviously, the replication of lagging strand proceeds from the replication fork towards the origin, *i.e.*, its direction is opposite to that of the leading strand.
9. After some more time of replication of the leading (3′ → 5′) strand, primase again initiates primer synthesis on the 5′ → 3′ (lagging) strand near the fork. The primer RNA thus formed provides the free 3′–OH for replication of the single-stranded region of the lagging strand.
10. Clearly, replication of the lagging (5′ → 3′) strand generates small polynucleotide fragments called '***Okazaki fragments***'. The replication of this strand is discontinuous in that it has to be initiated several times, and every times one Okazaki fragment is produced.
11. The RNA primer associated with the newly synthesized DNA strands/Okazaki fragments is most likely digested by DNA polymerase I in prokaryotes. This enzyme also catalyzes the filling of gaps so generated in the new strands. The Okazaki fragments (after the gap-filling by DNA polymerase I) are joined together by the enzyme polynucleotide ligase.
12. The above description pertains to the DNA replication progressing in one direction from the origin; the same events also occur in the opposite direction of the origin.

DNA Replication: Variations from Generalized Model-I

Strand displacement replication : In adenoviruses:

1. Begins at one end.
2. Only one strand is replicated.
3. The other strand separates as single strand, forms a ring, and
4. Replicates continuously.

- *Rolling Circle Replication*: In ϕx174:

1. A nick is induced at origin of circular chromosome.
2. The 3′–OH so produced is used as primer for DNA synthesis.
3. The 5′-end is progressively displaced as single strand.
4. When replication reaches origin, the single strand is nicked again.

24.6. VARIATIONS FROM THE GENERALIZED SCHEME

Variations in DNA replication occur in the manner of replication initiation, the scheme of replication fork movement and the timings of the replication of the two strands of the DNA molecule. These variations are represented in the following models of DNA replication: (1) strand displacement, (2) rolling circle and (3) D-loop mechanism of DNA replication.

24.6.1. Strand Displacement DNA Replication

In this scheme, replication begins at one end of a linear genome (= chromosome), *e.g.*, in adenoviruses. Only one of the two strands, the strand oriented in 3′→ 5′ direction, of the DNA duplex is replicated in a continuous manner. This produces a DNA duplex and a single strand (the 5′→ 3′ strand). The single strand base pairs at its ends to from a circle: replication of this strand now proceeds in a continuous manner (Fig. 24.7).

24.6.2. Rolling Circle Replication

Rolling circle replication occurs in many small bacterial and eukaryotic viruses, *e.g.*, in bactriophage ϕX174 (ϕ, Greek letter 'phi' here stands for phage). Phage protein A induces a nick, *i.e.*, break in a single strand, in one strand (the + strand) of the double-stranded chromosome. The nick occurs at the origin, and protein A binds to the 5′-end so produced. The 3′-end produced at the nick serves as primer and DNA synthesis proceeds from this end. As a result, the + strand is progressively displaced from the (–) strand beginning at the 5′-end. Replication produces a

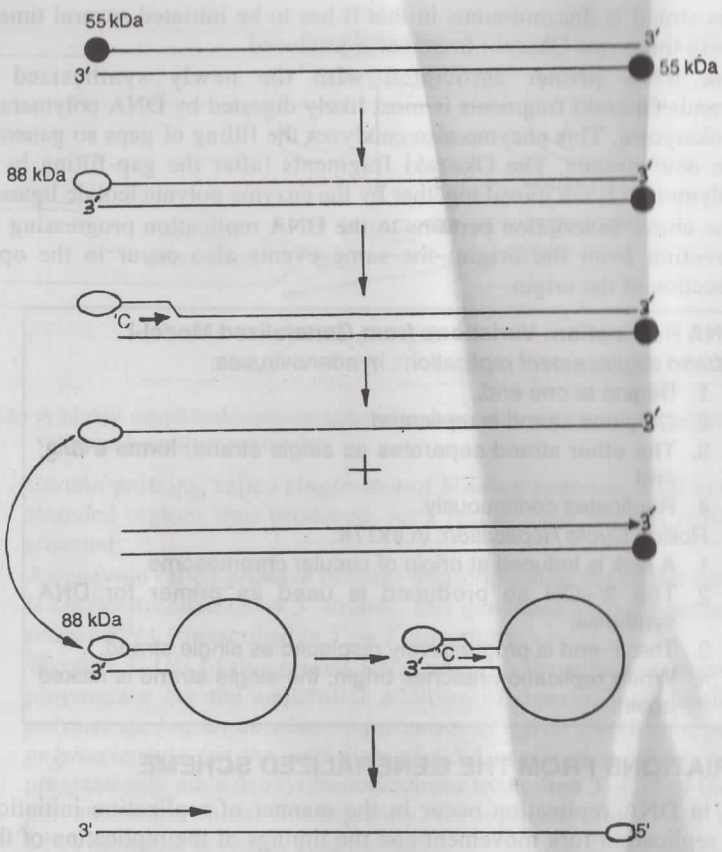


Fig. 24.7. Schematic representation of strand displacement mechanism of replication in adnovirus. The 88 kDa protein complex is replication complex.

single strand copy of the + strand, which functions as the viral genome. When a complete positive strand has been synthesized, and a new origin is produced in the (+) strand, protein A nicks in the origin to produce one single strand copy of the phage genome (Fig. 24.8).

24.6.3. Displacement Loop DNA Replication

This mode of replication occurs in mammalian mitochondria; the mitochondrial chromosome in circular. One strand of the chromosome is denoted as ***H strand***, while its complementary strand is named ***L strand***. Replication begins at a specific origin, but only one strand, the H strand, is replicated; the other strand (the *L* strand) of the chromosome is displaced forming a loop, called ***displacement loop*** or ***D loop***. When replication of (the H strand) has progressed up to about 2/3 of the chromosome, the replication of the displaced single strand (L strand) begins. This replication begins at a different origin and replication proceeds in the opposite direction. Thus both the strands of DNA are replicated in a continous manner, their replications begin at different origins, and replication of one strand (H strand) begins much earlier than that of the other (L strand). In this case, replication is not only

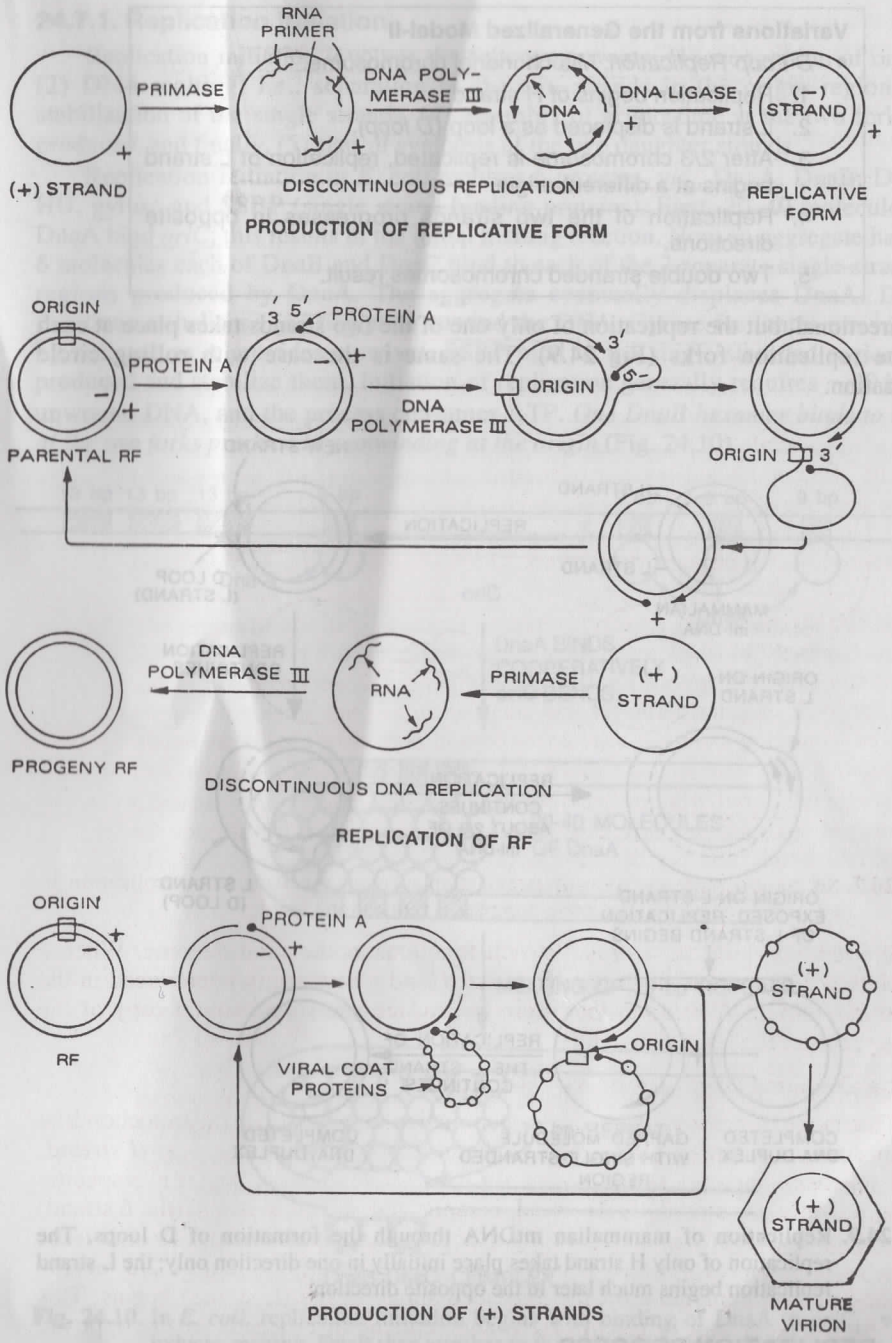


Fig. 24.8. Rolling circle replication of phage $\phi X174$ chromosome.

Variations from the Generalized Model-II

- *D Loop Replication*: Mitochondrial chromosome.
  1. Replication begins of H strand.
  2. L strand is displaced as a loop (*D loop*).
  3. After 2/3 chromosome is replicated, replication of L strand begins at a different origin.
  4. Replication of the two strands progresses in opposite directions.
  5. Two double stranded chromosomes result.

unidirectional, but the replication of only one of the two strands takes place at each of the replication forks (Fig 24.9). The same is the case with rolling circle replication.

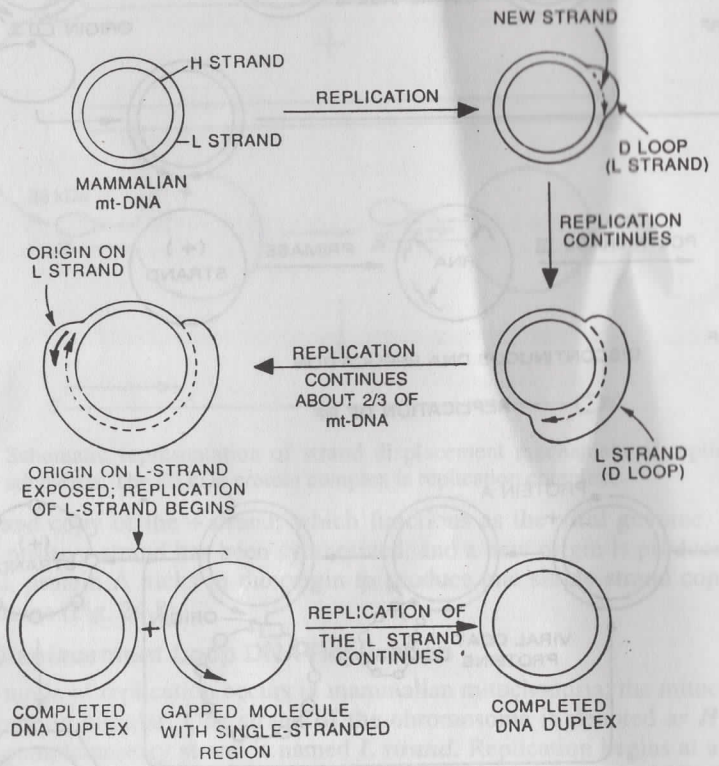


Fig. 24.9. Replication of mammalian mtDNA through the formation of D loops. The replication of only H strand takes place initially in one direction only; the L strand replication begins much later in the opposite direction.

24.7. REPLICATION PROCESS

DNA replication is a complex event; it can be divided into the following 3 major phases : (1) initiation, (2) elongation, and (3) termination.

24.7.1. Replication Initiation

Replication initiation involves the following events: (1) recognition of origin, (2) DNA melting, *i.e.*, separation of the two strands in the origin region, (3) stabilization of the single strands, (4) assembly of primosome at the two forks so produced, and finally, (5) start of synthesis of the two daughter strands.

Replication initiation in *E. coli* requires 6 proteins, *viz.*, DnaA, DnaB, DnaC, HU, gyrase and SSBP (single strand binding proteins). First, 20–40 molecules of DnaA bind *oriC*; this results in the initial melting reaction. Now an aggregate having 6 molecules each of DnaB and DnaC bind to each of the 3 separate single-stranded regions produced by DnaA. The aggregate eventually displaces DnaA. DnaB functions as helicase and begins to unwind the DNA. Gyrase facilitates unwinding by helicase as it provides a swivel. SSBP bind to the single-stranded regions so produced and stabilize them. Initiation of replication generally requires ~ 60 bp of unwound DNA, and the process consumes ATP. *One DnaB hexamer binds to each of the two forks produced by unwinding at the origin* (Fig. 24.10).

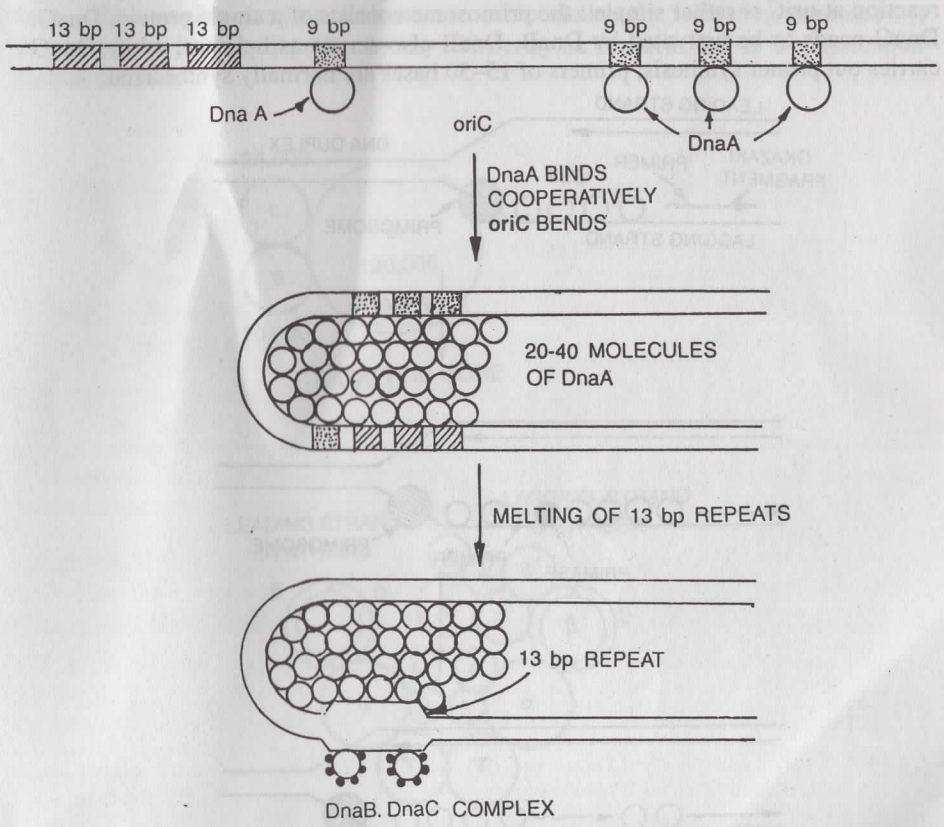


Fig. 24.10. In *E. coli,* replication initiation begins with binding of DnaA to *oriC,* which induces melting. DnaB then attaches to the potential forks and begins unwinding.

Once a replication fork is generated, ***primosome*** assembles at the origin, and initiates primer synthesis; this is called ***priming***. *Priming occurs only once and at*

Replication Initiation

In E. Coli at *oriC* :

1. DnaA (20-40 molecules) binds *oriC*; produces 3 separate single-strand regions
2. 6 molecules each of DnaB and DnaC bind each single-stranded region, and displace DnaA.
3. DnaB (helicase) unwinds DNA.
4. Gyrase facilitates unwinding of DNA.
5. SSBP stabilize single strands.
6. DnaB hexamer binds to each of two forks produced at origin.
7. Primosome assembles at *oriC*; DnaG is the primase; it synthesizes RNA primer
8. DnaG is activated by DnaB.
9. Priming of lagging strand (5′→ 3′ strand) occurs repeatedly.

the origin for the replication of the leading strand. But for replication of the lagging strand, priming occurs repeatedly at intervals of 1000 *to* 2000 *bases.* Priming reaction at *oriC* is rather simple : the primosome consists of a single protein, DnaG. DnaG needs to be activated by DnaB. DnaB also serves as helicase, while DnaG carries out primer synthesis; primers of 15–50 bases are normally synthesized.

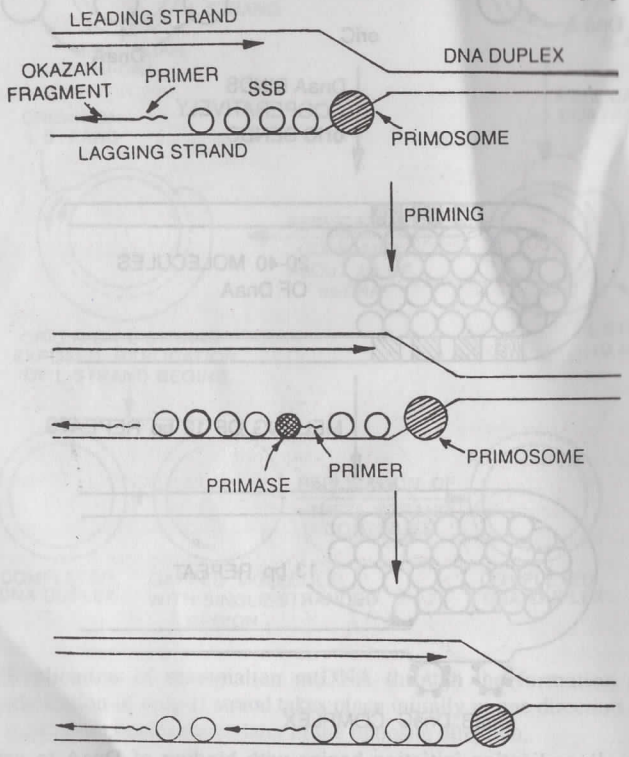


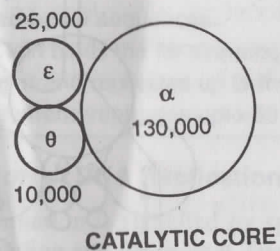
Fig. 24.11. A simplified schematic representation of the events involved in priming during replication of the lagging strand. Prisomosome may consist of simply DnaG at *oriC* or DnaG plus 5 other proteins in case of ϕX type replicons.

A different type of priming also occurs in *E. coli* at a different set of its replicons. This type of priming is called ϕ *X* type priming and the primosome consists of 6 proteins (PriA, PriB, PriC, DnaT, DnaB and DnaG). Dna*G* provides primase activity, while DnaB serves as helicase.

The replication fork proceeds in the 5′→ 3′ direction in relation to the lagging strand. The replication fork advances and generates a single–stranded region of the lagging strand bound to SSBP ahead of the primosome. The primosome moves along this single-stranded region. When the primosome reaches a site at which priming can occur, it synthesizes an RNA primer. This primer sponsors synthesis of a new Okazaki fragment (Fig. 24.11)

24.7.2. Primer Elangation (DNA Replication)

Once the primer has been synthesized, DNA synthesis is taken up by ***replisome,*** which is a complex of proteins. In *E. coli*, DNA replication activity is provided by DNA polymerase III component of replisome. Each *E. coli* cell has ~ 10 molecules of DNA polymerase III; most of these molecules are associated with replication forks. The complete enzyme, *holoenzyme*, molecule has the following subunits; α_2, θ_2, ε_2, γ, χ, ψ, δ, δ', τ_2, β_4 (Fig. 24.12) . The enzyme is assembled at the replication fork as follows.



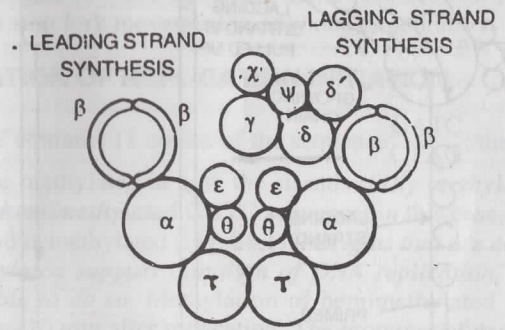


Fig. 24.12. A schematic representation of the organization of DNA polymerase III catalytic core and holoenzyme molecule.

1. First, the γ-δ complex (subunits γ δ δ′ χ ψ) and a pair of β subunit recognize the primed-template and bind to it.
2. They now attach to a catalytic core (∝ θ ∈ subunits).
3. Subunit τ now joins the complex. It brings two more β subunits and another catalytic core to the complex. This generates a DNA polymerase III holoenzyme.

Primer Elongation

- Replisome assembly (in *E. coli*);
  1. γ – δ complex and β subunit bind primed template.
  2. Now catalytic core (α θ ε submits) join.
  3. The τ subunit joins.
  4. Now two more β subunits and another catalytic core join.
- One catalytic core adds one nucleotide at a time to primer.
- The other catalytic core replicates the lagging strand.

According to one model, a single holoenzyme molecule functions at one replication fork. Each holoenzyme molecule has 2 catalytic cores; one catalytic core catalyzes the replication of leading strand, while the other catalyzes that of the lagging strand (Fig. 24.13). In the case of ***leading strand***, the catalytic core extends

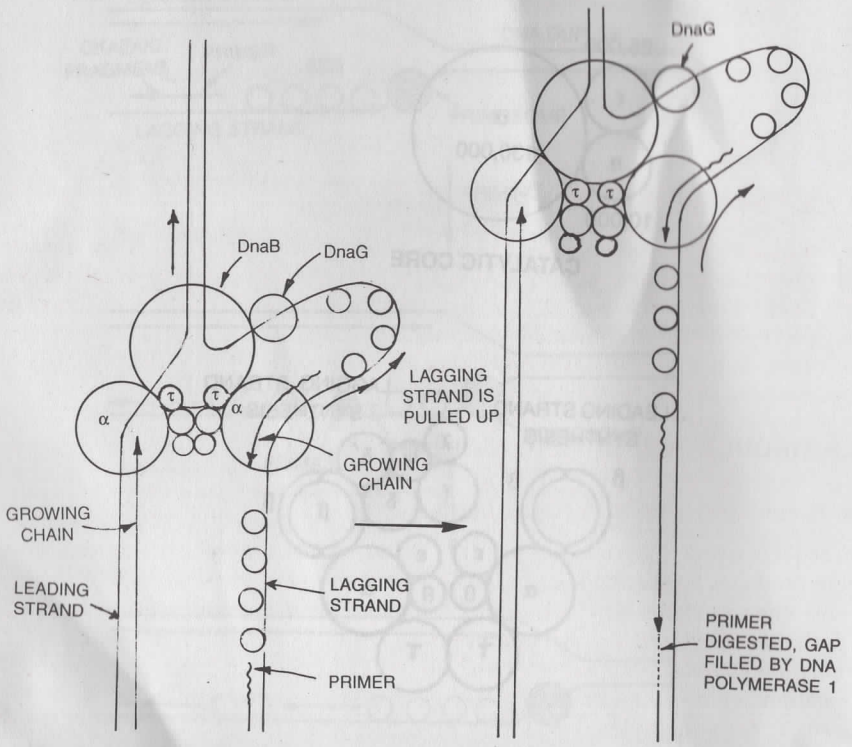


Fig. 24.13. Coordinated synthesis of leading and lagging strands by the same holoennzyme molecule of DNA polymerase III.

the primer one nucleotide at a time. DnaB progressively unwinds the duplex and the replication fork moves along.

Replication of the ***lagging strand*** will begin sometime later. When DnaB associated with the advancing fork reaches a site suitable for priming, it activates DnaG to synthesize a primer in the normal 5′ → 3′ direction, *i.e.*, moving from the fork toward the origin. When the primer become 10-14 bases long, the other catalytic core begins to elongate this primer in the 5′→ 3′ direction. The lagging strand, is in effect, pulled up by the replisome in the process of replication; it therefore, forms a progressively larger loop between the fork and the replisome (Fig. 24.13). When the replisome reaches the 5′-end of the primer of the previous Okazaki fragment, it stops replication and dissociates from the lagging strand.

Meanwhile DNAB continues to move forward with the replication fork. When it reaches the appropriate site, it again induces primer synthesis by DnaG and the events described above take place again.

In eukaryotes, two different enzymes are used to replicate the leading and the lagging strands. Leading strand is most likely replicated by DNA polymerase α, while replication of the lagging strand is due to DNA polymerase δ; this enzyme also provides primase activity.

Termination

In *E. coli*:

- Signalled by *ter* sequences.
- Tus protein binds the *ter* sequences; prevents fork movement.
- Leading strand replicates up to the *ter* sequence.
- Lagging strand replicates upto 50 bp prior to the *ter* sequence.

24.7.3. Termination of DNA Replication

In *E. coli*, termination is signalled by specific sequences called ***ter elements***, which serve as a binding site for protein Tus. Tus protein binds to *ter* element and stops DnaB from unwinding DNA. This stops the movement of the replication fork. The leading strand is replicated up to the *ter* element, while the lagging strand replication is stopped 50 –100 bp before the *ter* element. It is significant that Tus protein is able to stop fork movement in only one direction.

24.8. REGULATION OF REPLICATION INITIATION

The *E. coli, oriC* contains 11 copies of the sequence $\frac{\text{GATC}}{\text{CTAG}}$; the adenine residues of this sequence are methylated in both the strands (***fully methylated***). Replication of origin produces ***hemimethylated*** GATC sequence; in this case, the GATC sequence of only one strand is methylated (Fig. 24.11). *Origins that are either fully methylated or fully unmethylated support initiation of DNA replication, but hemimethylated origins are unable to do so.* Methylation of hemimethylated origins produced by replication begin ~10 min after replication. The promoter of gene *dnaA* also behaves like *oriC;* fully methylated promoter is active, while hemimethylated promoter is inactive. Transcription of *dnaA* is required for initiation of DNA replication as protein DnaA is the activator of replication initiation.

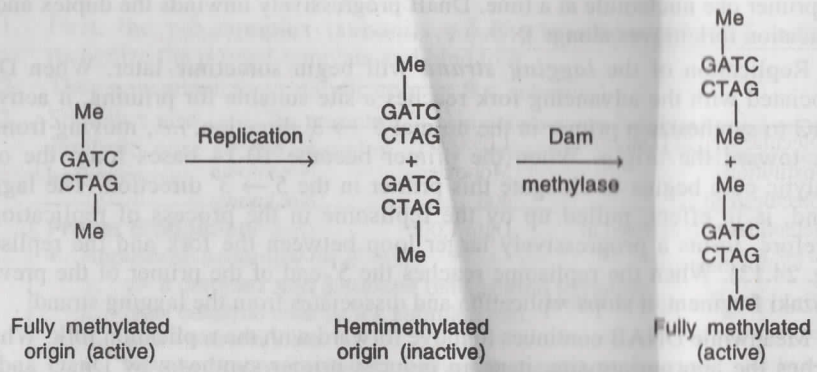


Fig. 24.14. Replication generates hemimethylated GATC sequences in origin; these are converted into fully methylated sequences by dam methylase.

Regulation of Initiation

- By the state of methylation of GATC sequence.
- *oriC* has 11 copies of GATC / CTAG.
- Fully methylated GATC supports replication initiation.
- Replication yields hemimethylated GATC sequences.
- Hemimethylated *oriC* attaches to cell membrance; this blocks initiation.
- Methylation occurs ~10 minutes after replication.

Hemimethylated *oriC* binds to bacterial membrane (fully methylated *oriC* does not bind). This binding to membrane prevents DnaA from binding to *oriC* and, thereby, replication initiation. Probably attachment of origin to membrane is important in the regulation of replication initiation.

24.9. PROTEINS INVOLVED IN DNA REPLICATION

In *E. coli*, at least two dozen gene products are involved in DNA replication. Most of these genes have been identified by isolating ***temperature-sensitive dna mutatants***; DNA in these mutants replicates normally at lower temperatures, but fails to do so at ~42°C. The important proteins/enzymes involved in DNA replication are briefly described below.

24.9.1. DNA Polymerase

DNA polymerase is the chief enzyme of DNA replication. DNA polymerase activity was discovered by Kornberg in 1956; this activity was due to DNA polymerase I. *E. coli* has two more enzymnes, DNA polymerase II and III, for DNA synthesis (Table. 24.1). All DNA polymerases require the following: (1) a template DNA strand, (2) a short primer (either RNA or DNA), and (3) a free 3′ –OH in the primer. They add one nucleotide at a time to the free 3′ –OH of the primer, and extend the primer chain in 5′→ 3′ direction.

TABLE 24.1
Properties of DNA polymerases I, II and III of *E. coli*

	DNA polymerase		
	I	***II***	***III***
Molecular weight (Daltons)	1,09,000	1,20,000	>2,50,000
Constitution	Monomer	Not known	Heteromultimer
Molecules/cell	400	Not known	10-20
Nucleotides polymerised at 37° C/min/molecule	Upto 1,000	Upto 50	Upto 15,000
Affinity for 5′-triphosphates of deoxyribonucleosides	Low	Low	High
Activities:			
5′ → 3′ polymerase	present	present	present
3′ → 5′ exonuclease	present	present	present
5′ → 3′ exonuclease	present	absent	absent
Functions in	DNA repair; excision of RNA primers	DNA repair	***DNA replication; it is the real replicase***
Mutant loci	*PolA*	*PolB*	*dnaE (polC), dnaN, dnaX, dnaQ, dnaT*

24.9.1.1. DNA Polymerase I

This enzyme provides the major part of activity in *E. coli.* It is *chiefly a DNA repair enzyme*, and is used for *in vitro* DNA replication. This enzyme has the following 3 activities.

1. The 5′→ 3′ polymerase activity is responsible for primer extension or DNA synthesis.
2. The 5′→ 3′ exonuclease activity is involved in excision of DNA strands during DNA repair; it removes ~ 10 bases at a time. An ***exonuclease*** digests nucleic acids (here DNA) from one end (it does not cut DNA internally).
3. The 3′→ 5′ exonuclease activity is responsible for proof-reading. In this case, only one nucleotide is removed at a time. DNA polymerase may scrutinized the bases that pair with the template strand during replication and allow only the correct bases to enter the chain; but it is not known if this really happens. *The polymerase action does commit errors in DNA synthesis.* DNA polymerase is known to scrutinize the new bases added to the growing chain and to delete or remove the wrong bases; this is called ***proof-reading***. Proof-reading activity is likely to reduce errors in replication by 100 – fold.

DNA polymerase I is encoded by gene *polA*, has a single polypeptide, and can initiate replication *in vitro* at a nick in a DNA duplex. It can be cleaved by proteolytic treatments into a large and a small fragments; the large fragment, called ***Klenow fragment***, lacks 5′ → 3′ exonuclease activity and is used for *in vitro* DNA replication.

24.9.1.2. DNA Polymerase II

This enzyme likely functions in DNA repair. It has 5′→ 3′ polymerase and 3′→ 5′ exonuclease activities, and uses as template only such DNA duplexes that have short gaps.

24.9.1.3. DNA Polymerase III

This enzyme is responsible for DNA replication *in vivo*. It has 5′→ 3′ polymerase and 3′→ 5′ exonuclease activities. It catalyzes DNA synthesis at very high rates, *e.g.*, 15,000 bases/min at 37°C. It is composed of several subunits as described in Section 7.2.

Proteins Involved in Replication

• DNA polymerases:	DNA synthesis
I :	Repair; removes primer, fills gaps
II :	Repair
III :	Replication
• Primase	Primer synthesis
• Helicase	DNA unwinding
• Gyrase	Swivel action
• DNA ligase	Seals nicks
• Endonucleases	Induces nicks
• SSBP	Stabilize single strands
• Pilot proteins	In viruses; determine if replication will occur.

A DNA polymerase molecule has the following 4 functional sites involved in polymerase activity (Fig. 24.15).

1. ***Template site*** binds the strand serving as template during replication.
2. ***Prime site*** binds to the primer used for DNA replication.
3. ***Primer terminus site*** binds only to such primers that have free 3′ –OH. This ensures that only primers having free 3′ – OH bind this site.
4. The ***nucleotide triphosphate site*** binds to the deoxynucleotide 5′-triphosphate that is complementary to the corresponding nucleotide of the template. It also catalyzes the formation of phosphodiester bond between the 5′ phosphate of this nucleotide and the 3′ –OH of the terminal primer nucleotide.

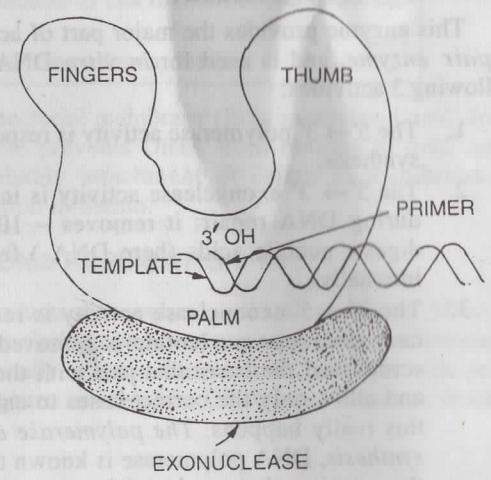


Fig. 24.15. Schematic representation of crystal structure of DNA polymerase belonging to *E. coli* pol I family. These subdomains create a DNA binding cleft, which is the polymerase catalytic domain; the active site lies in the palm. Proof-reading is provided by the exonuclease domain located below the palm.

[In addition, the polymerase molecule has (5) a 3′ → 5′ exonuclease site and (6) a 5′→ 3′ exonuclease site (in case of DNA polymerase I only)].

In case of eukaryotes, five different DNA polymerases are found (Table 24.2). ***DNA polymerase*** α replicates the leading strand, while ***DNA polymerase*** δ synthesizes the lagging strand. An accessory subunit of DNA polymerase α catalyzes priming in both the strands. ***DNA polymerases*** β and ε are probably nuclear DNA repair enzymes. ***DNA polymerase*** γ is found in mitochondria and chloroplasts. Surprisingly, only δ, ε and γ polymerases have 3′→ 5′ exonuclease activity (needed for proof-reading) and none of them has 5′→ 3′ exonuclease activity (needed for removal of primers during DNA replication).

TABLE 24.2
Different forms of DNA polymerase III of *E. coli*

Enzyme form	*Subunit organization*
Catalytic core	$\alpha\ \theta\ \varepsilon$
Pol III*	$\alpha_2\ \theta_2\ \varepsilon_2\ \tau_2$
Pol III′	$\alpha_2\ \theta_2\ \varepsilon_2\ \tau_2\ \gamma\ \delta\ \delta'\ \chi\ \psi$
Holoenzyme	$\alpha_2\ \theta_2\ \varepsilon_2\ \tau_2\ \gamma\ \delta\ \delta'\ \chi\ \psi\ \beta_4$

DNA Polymerase

- All enzymes need:
  1. A DNA template.
  2. A short primer.
  3. A free 3′ –OH in the primer.
- DNA polymerase I has the following activities.
  1. 5′→ 3′ polymerase.
  2. 3′→ 5′ exonulease (proof-reading)
  3. 5′→ 3′ exonuclease (repair).
- All enzymes have the following sites:
  1. Template site.
  2. Primer site.
  3. Primer terminus site.
  4. Nucleotide triphosphate site.
  5. 3′ → 5′ exonuclease site.
  6. 5′ → 3′ exonuclease site (DNA polymerase I only).

24.9.2. Primase

This enzyme activity catalyzes the synthesis of RNA primers to initiate DNA replication (Section 24.7.1). In *E. coli*, DnaG functions as primase. But in eukaryotes, a subunits of DNA polymerase α provides this function. There are, however, several other ways in which primers are produced, *e.g.*, the 3′–OH generated by a nick in the template DNA molecule (Section 24.6.2).

24.9.3. DNA Unwinding Proteins

These proteins are concerned with DNA unwinding during DNA replication, a brief description of them is given in Section 24.7.

24.9.4. Polynucleotide Ligase

DNA ligase or ***polynucleotide ligase*** catalyzes the formation of phosphodiester linkage between two immediate neighbour nucleotides of a DNA strand. Thus it seals the nicks remaining in a DNA strand either following DNA replication or DNA repair. However, this enzyme can not fill the gaps in DNA strands.

24.9.5. Endonucleases

An ***endonuclease*** produces an internal cut (single- or double-stranded) in a DNA molecule. But a ***restriction endonuclease*** produces cuts only at those sites that have a specific base sequence. During DNA replication, an endonuclease may induce a nick to initiate DNA replication (Section 24.6.2), or it may induce nicks to generate a swivel for DNA unwinding. Restriction endonucleases are required for DNA repair.

24.9.6. Pilot Proteins

Pilot proteins are produced by most viruses. The type of pilot proteins associated with viral genome determines whether the viral DNA will undergo replication or it would support transcription.

Replication of Genetic RNA

- *Direct replication* by RNA replicase: RNA duplex → RNA duplex
- *Replication via DNA*: RNA → DNA → RNA; cDNA produced by reverse transcriptase.

24.10. REPLICATION OF GENETIC RNA

Replication of genetic RNA may occur in one of the following two ways: (1) direct replication of RNA (RNA → RNA) and (2) RNA replication via DNA (RNA → DNA → RNA).

24.10.1. Direct Replication of RNA

In many viruses, *e.g.*, polio virus, influenza virus, double-stranded RNA phages, etc., RNA replication directly gives rise to RNA. This is called ***RNA-primed RNA synthesis***, and is catalyzed by ***RNA replicase*** or simply ***replicase***. Each RNA virus seems to produce a specific RNA replicase that recognizes only its own RNA for replication. The replication of double-stranded genetic RNA produces double-stranded progeny molecules. Single-stranded genetic RNA first becomes double-stranded. It then replicates to yield double-stranded molecules of which only the + strands are packed into virions.

24.10.2. RNA Replication *via* DNA

In case of some viruses, *e.g.*, certain tumour RNA (single-stranded) viruses, genetic RNA is first used as template by reverse transcriptase to direct the synthesis of its complementary DNA duplex (cDNA). The cDNA is then transcribed by RNA polymerase to yield genetic RNA.

24.11. DNA REPAIR

The damage to DNA may be grouped into the following two types: (1) single base changes and (2) changes in DNA structure. In case of ***single base changes***, a single

base of the DNA molecule is affected, there is no structural distortion and replication and transcription are not affected. An example of such a change is deamination of 5-methylcytosine that yields thymine; the resulting mismatched T·G. pair is resolved following DNA replication into one T·A and on G·C pair. In contrast, ***structural distortions*** generally interfere with replication and transcription. Examples of such changes are single strand nick, removal of a base, thymine dimers, etc.

Each cell possesses several repair systems to be able to deal with the various types of DNA damage. The repair systems may be grouped into the following 5 types: (1) direct repair, (2) excision repair, (3) mismatch repair, (4) tolerance systems and (5) retrieval systems.

DNA Repair

- *Direct repair*. Damage is undone, *e.g.*, thymine dimer resolved by Phr protein (in blue light).
- *Excision repair*. Three steps:
  1. *Incision* by an endonuclease.
  2. Excision and
  3. Synthesis; both by DNA polymerase I.
- *Very short patch repair*. Correction of base mismatch; few bases excised; enzymes MutL, MutS and MutY.
- *Short patch repair*. About 12 bases excised; endonuclease encoded by *uvrA, B* and *C; 99%* of excision repair activity.
- *Long patch repair*. For damages that block DNA replication; excision of 1,500 to >9,000 bases; enzymes Uvr*A, B* and *C* endonuclease, and DNA polymerase I.

24.11.1. Direct Repair of DNA

When the damage to DNA is reversed or simply removed, it is called ***direct repair***. For example, UV-rays induce thymine dimer formation in which the positions 4 and 5 of two thymine molecules located side by side in the same strand of DNA become covalently linked with each other (Fig. 24.16). In *E. coli*, protein product of gene *phr* specifically recognizes such dimers and breaks the covalent bonds between the two T residues. This protein binds to the dimer in dark, but requires energy from blue light for removal of the covalent bonds; for this reason, the phenomenon is known as ***photoreactivation***.

24.11.2. Excision Repair of DNA

In this repair pathway, the damaged segment of the DNA strand is removed and a new stretch is synthesized in its place. It involves the following steps (Fig. 24.17).

1. ***Recognition and Incision***. An endonuclease recognizes the damage and nicks the affected strand on either side of the damage.
2. ***Excision***. After this, an exonuclease removes the damaged section, generating a single-stranded region in the DNA duplex. In *E. coli*, the 5′→ 3′ exonuclease activity of DNA polymerase I carries out excision.
3. ***Synthesis***. In this step, the single-stranded region produced by excision is used as template for the synthesis of its complementary strand; this restores the duplex structure. In *E. coli*, DNA polymerase I carries out the synthesis step. The nick remaining after the synthesis step is sealed by DNA ligase.

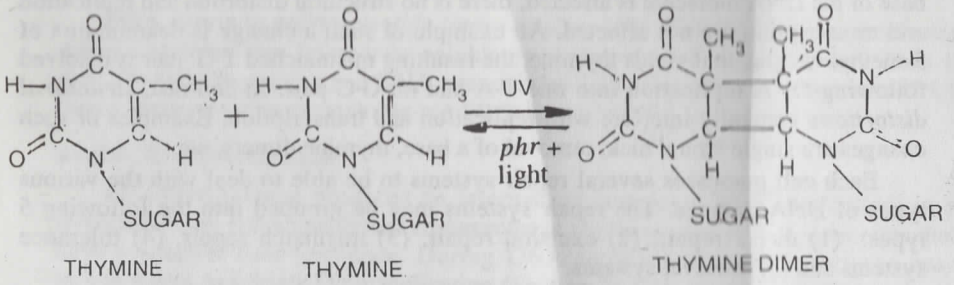


Fig. 24.16. Direct repair of DNA damage. Photolyase Phr breaks the two double bonds between thymine dimers produced by UV irradiation; this repair requires the presence of light.

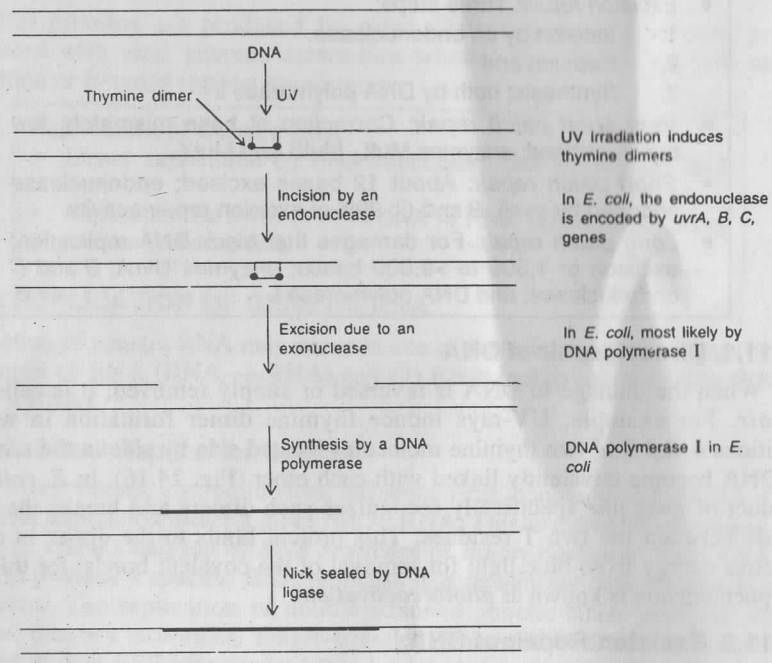


Fig. 24.17. A schematic representation of the events in excision-repair in *E. coli.*

The excision repair systems are grouped into the following three types based on the length of the DNA strand excised: (1) very short-patch repair, (2) short-patch repair and (3) long-patch repair (Table 24.3).

24.11.2.1. Very Short-Patch Repair

This system deals with repair of mismatches between specific bases. For example in *E. coil*, the mismatch errors produced during replication are corrected by the *dam* repair system. This repair system involves 3 enzymes, *viz.*, MutL, MutS and MutY, leads to the excision of few bases, and corrects the error present in the newly synthesized strand. The new 'strand' is recognized by the unmethylated state of A in

TABLE 24.3
A comparison among the different excision-repair systems

	Excision-repair system		
	Very short-patch	***Short-patch***	***Long-patch***
Type of damage repaired	Mismatches between specific bases	Pyrimidine dimers and other bulky adducts	Damages that block replication at damage site
Length of DNA strand excised	Few bases	~12 bases	1,500 - >9,000 bases
Enzymes involved	MutL, MutS, MutY	DNA polymerase I and an endonuclease (UvrA, B and C)	DNA polymerase I and an endonuclease (UrvrA, B and C)
Expression	Constitutive	Constitutive	Induced by DNA damage
Active on lesions	Any DNA strand; newly synthesized strand	Nonreplicating DNA	Located in the region of replication
Frequency of occurrence (in *E. coli*)	—	99% of excision repair events	1% of excision-repair events

the sequence $\begin{matrix}GATC\\CTAG\end{matrix}$. A in this sequence is methylated by Dam methylase (encoded by the gene *dam*).

24.11.2.2. Short-Parch Repair

In this systems ~12 bases long DNA strand is excised. For example, during excision repair of thymine dimers in *E. coli*, an endonuclease nicks the damaged strand 7 bases away from the 5′ side of the site of damage as well as 3 - 4 bases away from the 3′ side of the damage site. The endonuclease is produced by *E. coli* genes *uvrA, B* and *C*. Excision and repair steps are most likely the function of DNA polymerase I. It accounts for about 99% of excision repair activity in *E. coli*.

24.11.2.3. Long-Patch Repair

This repair system involves the excision of about 1,500 bases. (Sometimes upto >9,000 bases may be excised.) It represents ~1% of the excision repair activity in *E. coli*. The repair system is induced by a DNA damage that blocks replication. This system also involves the *uvr* genes and DNA polymerase I, and probably repairs the damage in the regions of replication forks.

24.11.3. Retrieval Systems in *E. coli*

These DNA repair systems are also known as ***recombination repair*** or ***post-replication repair*** systems. This systems operates when a structural distortion in DNA blocks replication at the damaged site. In such a situation, replication proceeds normally upto or close to the damaged site; DNA polymerase stops replication of the damaged strand, skips past the damaged site and reinitiates replication of this strand some distance away from the damaged site. However, the complementary strand of the DNA duplex is replicated normally. Thus replication of the damaged DNA molecule yields one normal DNA molecule, and one molecule having damage in one strand and a gap in the complementary strand; this gap can not be filled due to the

damage in the existing strand. This gap is filled by a single strand exchange with the normal progeny molecule; this exchange is catalyzed by the RecA protein. As a result of this exchange, the normal DNA molecule will now have a gap; this gap is readily filled by DNA polymerase I (Fig. 24.18).

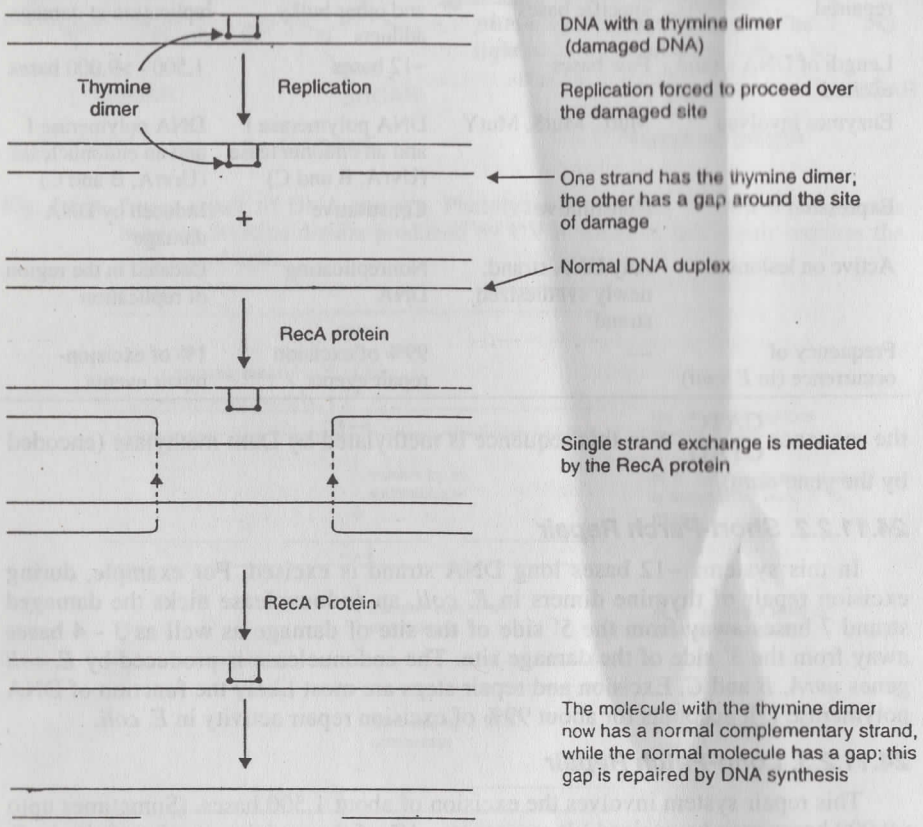


Fig. 24.18. A schematic representation of the recombination repair system in *E. coli.*

Retrieval system
- Replication of damaged strand does not occur at damage site.
- Single strand exchange between the two molecules produced by replication.
- Catalyzed by RecA protein.

24.11.4. Tolerance Systems

These systems deal with such damages that block normal replication of the damaged sites. They provide a means for the replication of damaged sites possibly with a high frequency of errors. These systems may be particularly important in the eukaryotes where the genome size is very large, and a complete repair of the damage is rather unlikely.

QUESTIONS

1. 'DNA replication is semiconservative and bidirectional.' Discuss the experimental evidence in favour of this statement.
2. DNA replication is a highly complex event. Briefly describe a generalised model of DNA replication, and explain the process of replication initiation.
3. Briefly explain the organization of replicon, and describe the different variations in the generalised scheme for DNA replication.
4. Explain DNA replication under the following heads:
   (*i*) replication initiation,
   (*ii*) primer elongation and
   (*iii*) replication termination.
5. Briefly describe the various proteins involved in DNA replication.
6. Briefly describe the various mechanisms of DNA repair.
7. Write short notes on the following:
   (*i*) semi-conservative replication, (*ii*) replicon,
   (*iii*) origin, (*iv*) terminus,
   (*v*) strand displacement replication, (*vi*) rolling circle replication,
   (*vii*) D loop replication, (*viii*) replication initiation,
   (*ix*) primer elongation, (*x*) replication termination,
   (*xi*) regulation of replication initiation, (*xii*) DNA polymerase,
   (*xiii*) DNA polymerase I, (*xiv*) primase,
   (*xv*) replication of genetic RNA, (*xvi*) DNA repair,
   (*xvii*) direct repair of DNA, (*xviii*) excision repair,
   (*xix*) retrieval systems in *E. coli.*

25

Organization of Genetic Material (Genes and Genomes)

25.1. INTRODUCTION

The genetic material of organisms is organized into structural units called ***chromosomes***. In prokaryotes, the entire genetic material of a cell is present in a single circular chromosome, while in eukaryotes it is distributed into several linear chromosomes. In addition, several other DNA sequences are also present either in cells or within genomes, *e.g.*, plasmids and transposable elements. Even the pattern of organization of genes also varies markedly. These aspects of the genetic material are considered in this chapter.

DNA

- *Unique DNA*:1 copy /genome:
  1. 8 – 50% of genome.
  2. produces > 80% of mRNA.
  3. *Genes*: constitutive, 10 – 15 thousand; luxury, 20 – 40 thousand.
- *Repetitive DNA* : Several to over one million copies /genome.
  1. 50 – 90% of genome.
  2. Types: (*i*) highly repetitive, (*ii*) moderately repetitive.

25.2. UNIQUE AND REPETITIVE DNA

In eukaryotes, DNA is classified into the following two categories: (1) unique DNA and (2) repetitive DNA. In prokaryotes, most of the DNA is unique. Some description of this aspect is given in Chapter 3 (Section 3.9.1).

25.2.1. Unique DNA

This class of DNA consists of those sequences that are present in a single copy per genome. In eukaryotes, the proportion of unique DNA ranges from only 8% of the genome in case of rye to ~50% in man. Virtually all the genes are represented in the unique DNA since more than 80% of the mRNA is produced by unique DNA. A rough estimate of the numbers of genes expressed in eukaryotes is as follows: in lower eukaryotes, *e.g.*, yeast, 4,000, and in somatic tissues of higher eukaryotes (plants and animals), 10,000-15,000. Thus the total amount of DNA represented in mRNA is only 1–2% of the total genome size. These estimates concern nonrepetitive DNA only. The mRNAs are termed as (1) abundant or (2) scarce. ***Abundant mRNAs*** are present in 1,000 – 10,000 copies per cell, and are of < 100 different types in any

cell. ***Scarce mRNAs***, on the other hand, are present in < 10 copies /cell, and consist of ~10,000 different sequences. In many cases, abundant mRNAs represent genes whose proteins are a major product of the concerned cell type.

In general, only ~10% of the mRNA sequences or species of a cell type are unique to it, *i.e.*, are present in only that cell type; a majority of the mRNA sequences is common to many, perhaps even all, cell types. Thus a set of about 10,000 genes is expressed in all the cell types of mammals; these genes are called ***house-keeping genes*** or ***constitutive genes***. The remaining genes represent specialized functions required in specific tissues or at specific stages of development; these are called ***luxury genes***. The total number of luxury genes may be between 20 and 40 thousand.

25.2.2. Repetitive DNA

This DNA fraction is made up of DNA sequences that are present in several to more than a million copies per genome. The proportion of repetitive DNA ranges from ~50% in human genome to ~90% in rye genome. Repetitive DNA is of the following two types: (1) highly repetitive DNA and (2) moderately repetitive DNA.

Highly Repetitive DNA

- > 10^5 copies/genome.
- Very short sequences.
- Sat-DNA differs in density from the rest of genomic DNA.
- Sat-DNA located in heterochromatic regions, *e.g.*, telomere, centromere; not transcribed.
- Minisatellites (VNTRs), are tandom repeats of 2–8 bp; highly variable.

25.2.2.1. Highly Repetitive DNA

These sequences are present in >10^5 to upto several million copies per genome. They usually consists of very short nucleotide sequences, but sometimes the sequences may be several hundred nucleotides long. Some highly repetitives sequences differ in density from the rest of chromosomal DNA; they form a separate minor band on $CsCl_2$ density gradient centrifugation and are called ***satellite DNA***. Satellite DNA occurs in almost all eukaryotes. It may be heavier or lighter than the ***main band DNA***, *i.e.*, the remaining chromosomal DNA. *Satellite DNAs are located in the heterochromatic regions of chromosomes, usually in centromeric regions.* In *Drosophila*, the three major satellite sequences are ACAAACT, ATAAACT and A C A A A T T. But mammalian satellites consist of hierarchical repeats in which variants of a small sequence are repeated to generate a larger sequence. Satellites may continually be generated and lost from genomes; presumably they are not under high selection pressure.

Sat-DNA sequences may have some structural function in the chromosome. In any case, they are not transcribed. Some sequences that resemble satellite DNA consist of tandem repeats of a short sequence of 2 to 8 bp, the repeats being identical or almost identical. Variation in this type of DNA arises due mainly to differences in the numbers of repeat units, which is highly variable. These DNAs are called ***minisatellites*** or ***VNTRs*** (variable number of tandem repeats) and are used for DNA finger-printing and genome mapping.

Moderately Repetitive DNA
- 2 – <10^5 copies/genome.
- 20 - 80% of genome.
- Sequences form distinct families.
- Genes for histones, rRNA, ribosomal proteins and several types of transposable elements.

25.2.2.2. Moderately Repetitive DNA

This class of DNA consists of a variety of sequence families that are present in two to less than 10^5 copies per genome. It makes up ~20 - 80%, of chromosomal DNA in eukaryotes. In a ***family***, the repeated sequences are not identical but are related to each other; the degree of relationship varies among the members of a family. *The members within a family hybridizes with each other, while those belonging to different families do not.* The members of these families are usually distributed throughout the genome in a more or less regular way with longer stretches of unique DNA separating them. No structural or functional significance can be assigned to this class of DNA as a group.

Moderately repetitive DNA includes genes for histones, ribosomal RNA and ribosomal proteins, and several types of transposable elements. But only a small proportion of genes is present in this class of DNA: ~10% of mRNA corresponds to moderately repetitive DNA, while 80% or more of mRNAs represent unique or nonrepetitive sequences.

Gene Organization
- *Uninterrupted Genes*: Coding region continuous; in eubacteria.
- *Overlapping Genes*: Two or more genes share the same DNA sequence; reading frames same or different; *e.g.*, $\phi X174$.
- *Interrupted Genes*: Coding region interrupted by introns; in eukaryotes (almost all genes); introns removed from RNA during splicing; uninterrupted genes evolved from interrupted genes.
- *Alternative Genes*: Genes whose RNA transcript is spliced in two or more different ways to yield 2 or more distinct proteins.
- *Antibody Gene Segments*: Functional gene assembled from distinct segments; *e.g.*, light chain gene from segment *V, J* and *C*. Yields very large number of genes from relatively few segments.
- *Pseudogenes*: Nonfunctional copies of genes; due to mutation following duplication; also from mRNA (processed pseudogenes).

25.3. ORGANIZATION OF GENES

Proteins represent those regions of genes that code for amino acids; these regions are called ***coding regions***. The coding region has translation initiation codon at its 5′-end in mRNA and a termination or nonsense codon at its 3′-end. In addition to the coding region, a gene must have sequences that specify various functions during transcription, and translation (see, Chapters 29, 30). In this consideration, we take

into account the organization of coding region only. Coding regions of genes may show different organizations as follows: (1) uninterrupted genes, (2) overlapping genes, (3) interrupted genes, (4) alternative genes, (5) gene segments and (6) pseudogenes.

25.3.1. Uninterrupted Genes

In case of ***uninterrupted genes***, the coding region is continuous, and all the bases/base sequences present in the gene between the initiation and termination codons are also present in the mRNA produced by this gene. Almost all the genes present in prokaryotes are uninterrupted. In general, genes in prokaryotes are organized into functional groups called operons; the various features of operons are considered in Chapter 31. Virtually all the genes in eukaryotes are functional units.

25.3.2. Overlapping Genes

Generally, neighboring genes are separated by a variable length of spacer DNA, and they are non-overlapping (Fig. 25.1). In such cases, a single DNA sequence is represented in only one protein. But in many viruses, *e.g.*, phage MS 2 phage $\phi X174$ and SV40, boundaries of some gene overlap either partially or fully; such genes are

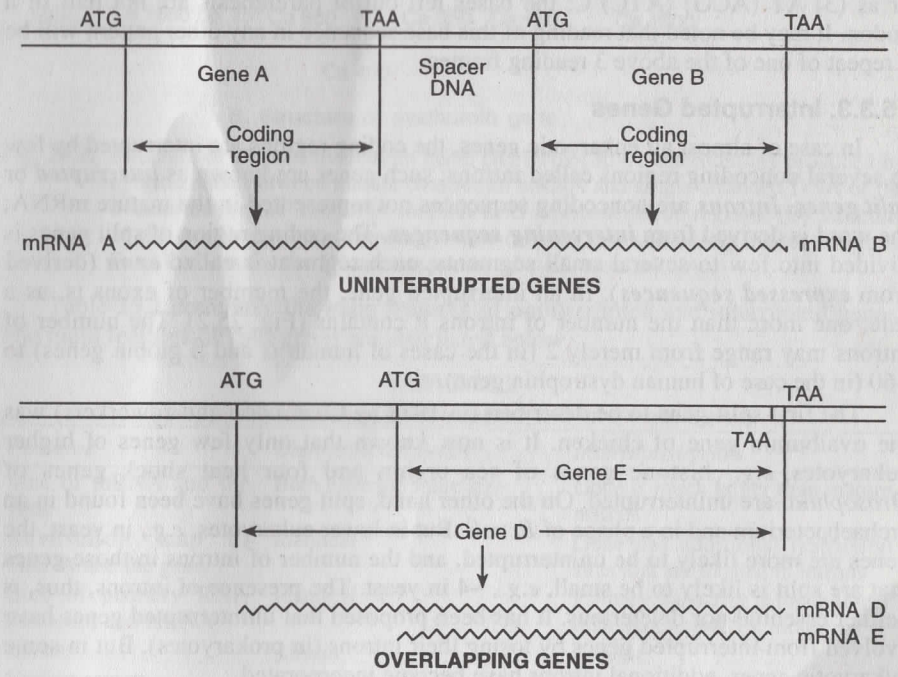


Fig. 25.1. Organization of uninterrupted (nonoverlapping) and overlapping genes (genes D and E of ϕX174). ATG (=AUG), initiation codon, and TAA (=UAA), termination codon.

called ***overlapping genes***. Thus in case of overlapping genes, a single DNA sequence is represented in two or more different proteins; this can happen in the following two ways.

1. In some cases, a larger and a smaller protein are synthesized from the same gene, and the reading frames for both are the same. The initiation of translation for the smaller protein takes place at an alternative initiation codon located within this gene. Both the polypeptides, however, terminate at the same site. Thus the smaller protein is, in fact, a portion of the larger protein produced by the gene.
2. In some other cases, the DNA sequence shared by the two genes (the overlapping sequence) is read in two different reading frames. As a result, the amino acid sequences encoded by the overlapping region are quite different in the two proteins. For example, gene *E* of $\phi X174$ is located within gene *D*. Gene *E* begins after a little more than one-third length of gene *D* and ends about 2 codons before gene *D*. The reading frame of gene *E*, however, is offset to the right by one nucleotide from that of gene *D*. A similar situation occurs in two other genes of this phage.

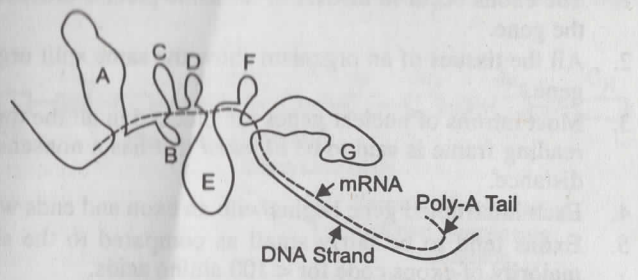
Reading frame of a gene refers to the sets of 3 bases forming the sequence of codons during translation. Since each codon has 3 bases, a given base can be read in 3 alternative frames. For example, base sequence ATACGATCC can be read as (1) (ATA) (CGA) (TCC) [each bracket denotes one codon], as (2) A (TAC) (GAT) CC, or as (3) AT (ACG) (ATC) C; the bases left out of parenthesis are not part of a codon. It may be noted that reading of this base sequence in any other pattern will be a repeat of one of the above 3 reading frames.

25.3.3. Interrupted Genes

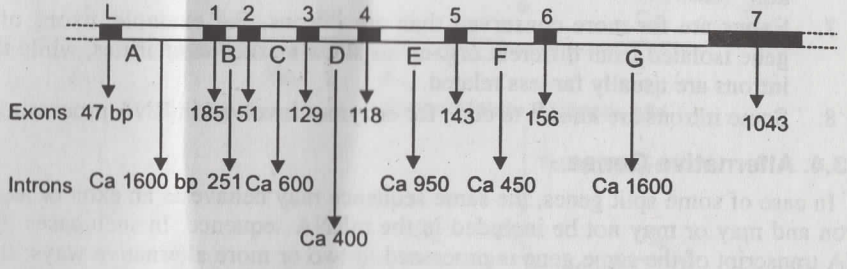
In case of almost all eukaryotic genes, the coding regions are interrupted by few to several noncoding regions called introns; such genes are known as ***interrupted*** or ***split genes***. ***Introns*** are noncoding sequences not represented in the mature mRNA; the word is derived from ***intervening sequences***. The coding region of split genes is divided into few to several small segments; each segment is called ***exon*** (derived from ***expressed sequences***). In an interrupted gene, the member of exons is, as a rule, one more than the number of introns it contains (Fig. 25.2). The number of introns may range from merely 2 (in the cases of human α and β globin genes) to >60 (in the case of human dystrophin gene).

The first split gene to be described (in 1977 by Chambdon and coworkers) was the ovalbumin gene of chicken. It is now known that only few genes of higher eukaryotes, *e.g.*, histone genes of sea urchin and four heat shock genes of *Drosophila*, are uninterrupted. On the other hand, split genes have been found in an archaebacterium and in a phage of *E. coli*. But in lower eukaryotes, *e.g.*, in yeast, the genes are more likely to be uninterrupted, and the number of introns in those genes that are split is likely to be small, *e.g.*, ~4 in yeast. The presence of introns, thus, is neither essential nor deleterious. It has been proposed that uninterrupted genes have evolved from interrupted genes by losing their introns (in prokaryotes). But in some eukaryotic genes, additional introns have become incorporated.

Transcription of interrupted genes yields RNA that contains both intron and exon sequences; this RNA (called ***primary transcript***) can not function as mRNA due to the presence of intron sequences. The primary transcripts is processed to precisely remove the intron sequences and to join the exon sequences in precisely the same order as before; this process is called ***splicing***, and it generates functional mRNA molecules.



A. DNA : RNA Heteroduplex



B. Structure of ovalbumin gene

Fig. 25.2. Structure of chicken ovalbumin gene. (A) DNA : RNA heteroduplex produced by hybridization between mature ovalbumin mRNA and single-stranded DNA of the ovalbumin region. The DNA loops A to G represent intron regions. (B) Structure of the ovalbumin gene showing intron (unsheaded) and exon (shaded) regions. Exon L codes for the 5′ leader sequence of ovalbumin mRNA.

Table 25.1 compares the main features of uninterrupted and interrupted genes. The main features of split genes are also listed below.

TABLE 25.1
A comparison between uninterrupted and interrupted genes

Features	*Uninterrupted genes*	*Interrupted genes*
Size of mRNA and the gene	Both have the same size	Gene is much larger than mature mRNA
Restriction map of mRNA and the gene	Identical	Gene has additional restriction sites that are located in introns
Size of primary RNA transcript and mRNA	Identical	Primary transcript is much larger than mature mRNA molecule
Introns	Absent	Present
Coding region	Continuous	Interrupted by introns
RNA splicing	Absent	Essential to generate mature mRNA
Occurrence	Prokaryotes; some eukaryotic genes; all genes of eubacteria	Eukaryotes; present in an archebacterium and a phage of *E. coli*
Number of proteins specified	One protein encoded by one gene	One DNA sequence may code for multiple proteins

1. The exons occur in mRNA in the same precise order in which they occur in the gene.
2. All the tissues of an organism show the same split organization for a given gene.
3. Most introns of nuclear genes are blocked in all the three reading frames. A reading frame is said to be ***blocked*** if it has a nonsense codon after a short distance.
4. Each interrupted gene begins with an exon and ends with an exon.
5. Exons tend to be fairly small as compared to the size of the gene. The majority of exons code for < 100 amino acids.
6. Introns are longer than exons; intron size ranges from 7,200 bp to 50–60 kb.
7. Exons are far more conserved than are introns. For example, exons of a gene isolated from different organisms show marked similarities, while the introns are usually far less related.
8. Some introns are known to code for enzymes involved in RNA processing.

25.3.4. Alternative Genes

In case of some split genes, the same sequence may behave as an exon or as an intron and may or may not be included in the mRNA sequence. In such cases, the RNA transcript of the same gene is processed in two or more alternative ways; this is called ***alternative splicing***. Alternative splicing yields two or more types of mRNA molecules from a single gene; these mRNA share a common sequence but differ for a part of their sequence. In a simple situation, exon 1 (E_1) may be separated from E_2 by a sequence I_1. I_1 may function as either intron or exon. When I_1 functions as an intron, the mRNA sequence will be E_1–E_2. But when I_1 functions as an exon, the mRNA will be E_1–I_1–E_2. In this way, sequences E_1 and E_2 will be represented in two different proteins. Such genes are described as ***alternative genes***. Sometimes the alternative forms by mRNAs may be produced simultaneously in the same cell. Alternatively, they may be produced either in different cell types or under different conditions in the same cell.

Clearly, alternative genes are like typical interrupted genes in their organization. But their RNA transcripts are processed in two or more alternative ways to produce two or more mRNAs with some differences in their sequence.

25.3.5. Antibody Gene Segments

Genes coding for antibodies have a peculiar organization. The functional gene is produced by joining together of 3 or 4 different gene segments in a precise order (Fig. 25.3). Most of these gene segments are present in more than one form, and some of them, *e.g.*, segment *CH*, are interrupted. The different gene segments are separated from each other by noncoding spacer sequences. These spaces contain at their ends signal sequences that are involved in the correct joining of these segments through a process of recombination.

Each antibody molecule is composed of two types of chains called ***light*** (smaller) and ***heavy*** (longer) ***chains***. In case of light chains, the variable region of the chain is encoded together by segments *V* and *J*, while the constant region of the chain is specified by the segment *C*. But in case of heavy chain, the variable region

is coded for by segments *V*, *D* and *J*, while the constant region is encoded by *C* segment.

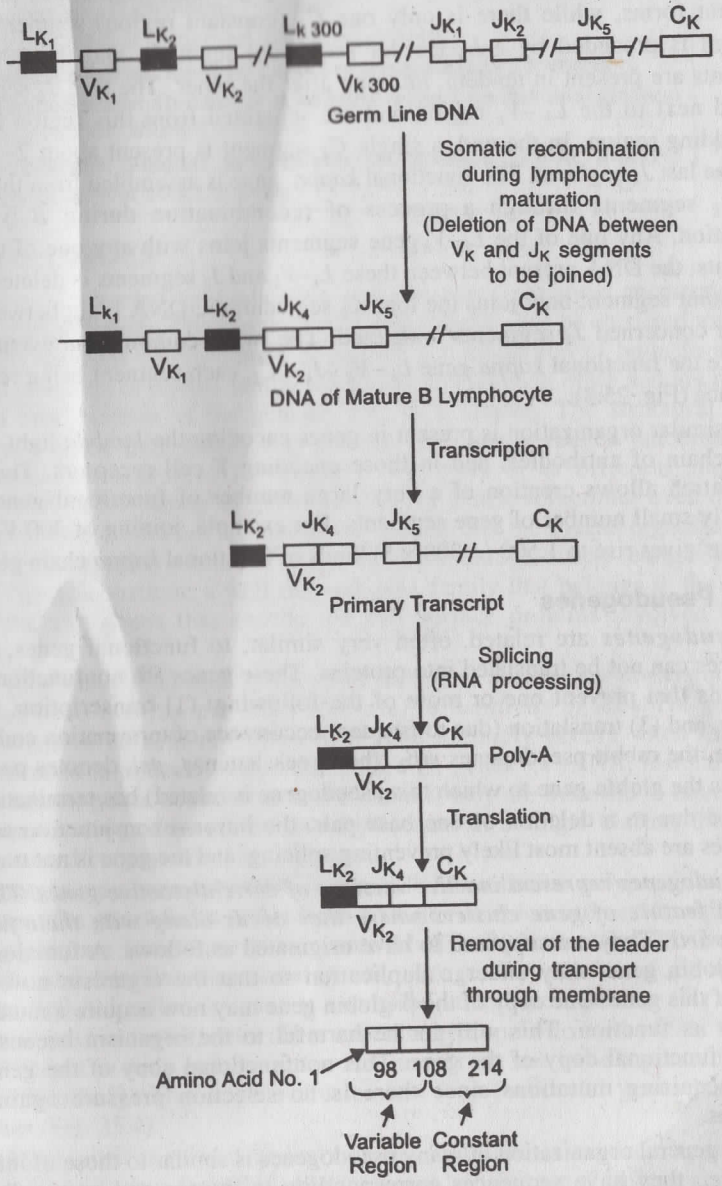


Fig. 25.3. Organisation of the segments coding for the kappa light chain of antibodies. Somatic recombination of the V_K and J_K segments produces the functional genes for different species of the kappa chain. Any of the 300 $L_K - V_K$ segment combinations may join with any of the five J_K segments.

In case of *kappa* light chain, the *V* segment is present in ~300 different V_k (variable region of *kappa* chain) forms, the J_k (joining) segment is present in 5 different forms, while there is only one C_k (constant region) segment. Each V_k segment is preceded by a L_k (leader sequence) segment, and the 300 $L_k - V_k$ segments are present in tandem, *i.e.*, one after the other. The five J_k segments are located next to the $L_k - V_k$ region and are separated from this region by a large noncodding region. In the end, a single C_k segment is present about 2–3 kb away from the last J_k segment. The functional *kappa* gene is assembled from the L_k–V_k, J_k and C_k segments through a process of recombination during *B* lymphocyte maturation. Any one of the L_k–V_k gene segments joins with any one of the five J_k segments; the DNA present between these L_k–V_k and J_k segments is deleted. The L_k–V_k -J_k joint segment now joins the lone C_k segment; the DNA lying between the C_k and the concerned J_k segments is deleted. The two recombination events thereby generate the functional *kappa* gene $L_k - V_k - J_k - C_k$, each segment being represented only once (Fig. 25.3).

A similar organization is present in genes encoding the *lambda* light chain and heavy chain of antibodies, and in those encoding *T* cell receptors. This kind of originzation allows creation of a very large number of functional genes from a relatively small number of gene segments. For example, joining of 300 V_k and 5 J_k segments gives rise to 1,500 (= 300 × 5) kinds of functional *kappa* chain genes.

25.3.6. Pseudogenes

Pseudogenes are related, often very similar, to functional genes, but their sequences can not be translated into proteins. These genes are nonfunctional due to mutations that prevent one or more of the following: (1) transcription, (2) RNA splicing, and (3) translation (due to frequent occurrence of termination codons). For example, the rabbit pseudogenes $\psi\beta_2$ (the Greek letter ψ, *psi*, denotes pseudogene and β_2 is the globin gene to which this pseudogene is related) has termination codon produced due to a deletion of one base pair, the intron–exon junction consensus sequences are absent most likely preventing splicing, and the gene is not transcribed.

Pseudogenes represent inactive versions of currently active genes. They are a common feature of gene clusters where they occur along with their functional counterparts. They are supposed to have originated as follows. A functional gene, say β-globin gene, may undergo duplication so that the organism now has two copies of this gene. One copy of the β-globin gene may now acquire a mutation that prevents its function. This will not be harmful to the organism because it still has one functional copy of the gene. This nonfunctional copy of the gene would go on acquiring mutations since there is no selection pressure against these mutations.

The general organization of many pseudogenes is similar to those of interrupted genes, *i.e.*, they have sequences corresponding to introns and exons. But many pseudogenes lack introns and resemble mature mRNAs produced by their active counterparts; they are called ***processed pseudogenes***. Processed pseudogenes are presumed to have originated by reverse transcription of the concerned mRNAs; the cDNA so produced then became inserted into the genome. An example of a processed pseudogene is the mouse ψα3 globin gene.

Gene Families

- Genes derived from a single ancestral gene by mutation following duplication.
- Clustered or dispersed (by translocation).
- *Gene Cluster*: identical copies of a gene located in tandem.
- *Superfamily*: Genes/gene families having similar organization but rather different functions.
- *Example:* β-globin gene cluster has genes ε, G_γ, A_γ, δ and β spread over 50 kb of DNA.

25.4. GENE FAMILIES

A set of genes descended by duplication and variation from some ancestral genes constitutes a ***gene family***. The members of a gene family may be clustered together, dispersed on different chromosomes or present a combination of both. In a ***gene cluster***, the members of a gene family are identical with each other and are present together at one location in the genone, *i.e.*, in a cluster. The structural genes belonging to a single gene family may perform the same or related functions; they may, however, be expressed in different cell types or at different stages of development. A ***super-family of genes*** consists of genes, sometimes even discrete gene families, whose roles are apparently not related. But the overall organization of these genes is noticeably similar. For example, immunoglobulins (comprising the antibodies) genes constitute a well defined gene family that belongs to the super-family containing genes that encode the cell surface proteins involved in cell adhesion.

Some gene families consist of identical members; such genes always occur in clusters and have two (on the lower extremé) to hundreds of identical genes in tandem. Extensive tandem repetition of a gene normally occurs when the gene product is needed in unusually large amounts, *e.g.*, genes for rRNA, histone genes, etc. When related genes occur at several locations, they are believed to have arisen through translocation of members initially located in a single cluster; the genes usually become divergent after they disperse.

The adult human haemoglobin is composed of two similar polypeptides, *viz.*, α and β chains. The genes for α globin lie in a cluster on chromosome 16, while those for the β globin are located in chromosome 11. The β cluster (Fig. 25.4) extends over 50 kb, and has five functional genes, *viz.*, ε, *Gγ*, *Aγ*, δ and β, and one pseudogene ($\psi\beta_1$). Epsilon (ε) gene is expressed in embryos, gamma (*Gγ* and *Aγ*) chain is expressed in foetus, delta (δ) is expressed just prior to birth, while beta (β) expression occurs in early embryo and is almost the only one to be expressed during adulthood. Similarly. The α gene cluster has four functional genes and three pseudogenes (Fig. 25.4).

Multiple copies of related genes may be organized in tandem repeats of a repeating unit, *e.g.*, genes for *r*RNA, histone genes, etc. In case of histone genes, one repeat unit consists of one copy each of the five histone genes, *viz.*, genes for H_1, H_2a, H_2b, H_3 and H_4 (Fig. 25.5). Each gene in the repeating unit is separated from its neighbour by a nontranscribed spacer region. The general organization of the histone clusters remains the same in organisms ranging from fruit fly, through sea urchain to mammals. But the gene clusters in different organisms differ in respect of

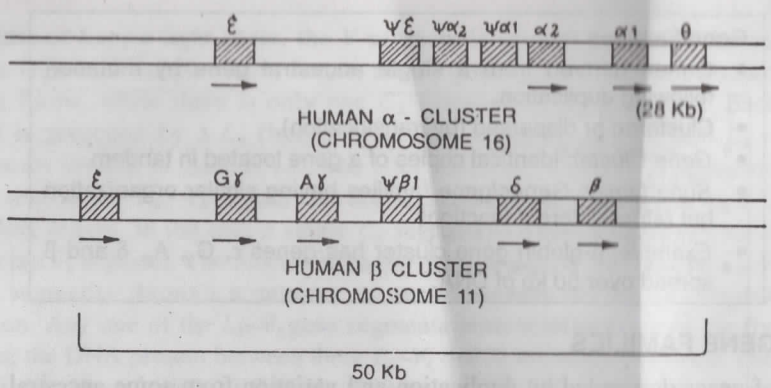


Fig. 25.4. α-like and β-like gene clusters containing α and β pseudogenes (represented by ψ) in human chromosomes 16 and 11, respectively. The genes are shaded dark; active genes have arrows below them (the direction of arrows indicates the orientation of the reading frame), while the pseudogenes are indicated by the prefix ψ and the lack of arrows below them.

(1) the number of repeat units present in the cluster, (2) the size of repeat units and (3) the orientation of member genes within the repeat units (Fig. 25.5). It may be pointed out that histone genes do not have introns. Similar tandem repeats of repeating units are found in case of 5S RNA genes, tRNA genes, snRNA (small nuclear *RNA*) genes, etc.

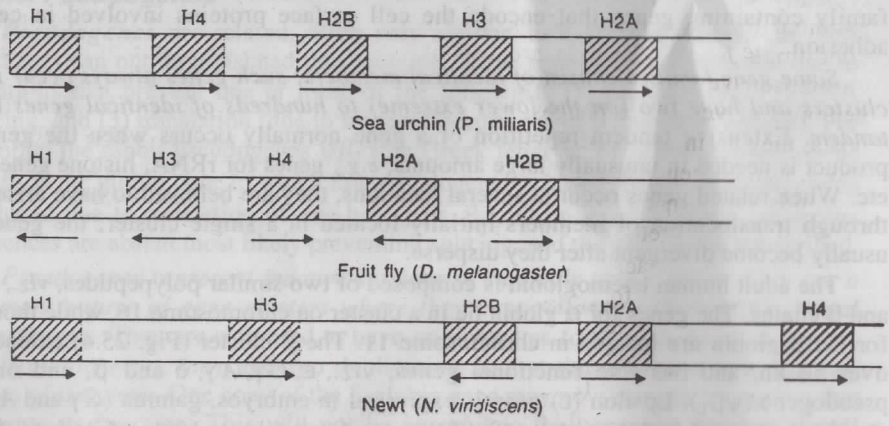


Fig. 25.5. Organization of the repeat unit of histone gene clustes in certain organism.

Members of a gene family are believed to have originated from a single ancestral gene. The ancestral gene would have undergone duplication to give rise to two copies of the gene. The two copies of the gene would accumulate mutations and diverge, *i.e.*, become different, from each other. One of the two copies will be under selection pressure to retain the original function. But the other copy would be freed from selection pressure and would become gradually more and more divergent. In the early stages of divergence, the products of two copies will appear as variants with the same function, *e.g.*, β and γ globin genes. But in due course, they will diverge to acquire different functions, *e.g.*, α and β globin genes. The two or more

copies of a gene may undergo duplication again; this will increase the number of members in the gene family. Thus duplication is the key to the origin of gene clusters.

Plasmid

- Circular DNA molecule inherited stable in extra chromosomal state.
- *Episome*: Replicons that may become integrated into or remain independent of bacterial chromosome.
- Type: 1. *F plasmids*: Conjugation.
  2. R. plasmids: Antibiotic resistance.
  3. Col plasmids: Colicin production.

  Also, conjugative and non-conjugative.
- *Copy number*: (*i*) Single copy (stringent control), or (2) multicopy (relaxed control).
- *Compatibility group*: Plasmid of the same group can not reside in the same cell; they have the same mechanism of replication control.

25.5. PLASMIDS AND EPISOMES

Bacterial cells may contain, in addition to their circular chromosomes, some circular minichromosomes called plasmids. ***Plasmids*** are replicons that are stably inherited in an extrachromosomal state. Plasmids do get eliminated but at a very low rate, typically $<10^{-7}$/cell. Most but not all plasmids are dispensable as they are not essential for the survival of bacterial cells, except under certain environmental conditions. Some plasmids are capable of becoming integrated into bacterial chromosomes; they are known as ***episomes***. Some other DNA sequences, *e.g.*, genomes of temperate phages, can also integrate into the chromosomes of their hosts; they are also called episomes. Thus episomes can exist in either of the two states, *viz.*, an independent state or in a state of integration in the host cell genome. Plasmids vary in size from those having only 3 genes to those containing several hundred genes. Bacterial cells may contain zero to several (upto 11) different plasmids.

25.5.1. Types of Plasmids

There are several types of bacterial plasmids of which the following 3 types have been extensively studied.

1. ***F plasmids*** are responsible for bacterial conjugation; they carry genes for the development of *F* pili and for conjugation.
2. ***R plasmids*** carry genes for resistance to antibiotics and antibacterial drugs; earlier they were called RTF (resistance transfer factors).
3. ***Col plasmids*** code for colicins, which are proteins that kill sensitive *E. coli* cells.

A plasmid may be either (1) conjugative or (2) nonconjugative. ***Conjugative*** or ***transmissible plasmids*** are capable to sponsoring conjugation, *e.g.*, F and F′ plasmids, many R plasmids and some Col plasmids. These plasmids rapidly spread among bacterial cells of a population through the process of conjugation; this transfer can occur across species as well, *e.g.*, transfer of *E. coli* R plasmids to

Salmonella, Hemophilus, Shigella, etc. In contrast, ***nonconjugative*** or ***nontransmissible plasmids*** do not mediate DNA transfer through conjugation, *e.g.*, many R and most Col plasmids.

25.5.2. Copy Number

Each plasmid is maintained in bacterial cells at a characteristic copy number. The regulation of plasmid copy number is based on plasmid replication control system. The plasmids are grouped into the following two types on the basis of their number /cell.

25.5.2.1. Single Copy Plasmids

These plasmids are maintained in one copy per cell. The replication control of such plasmids is the same as that of bacterial chromosome, and they replicate and segregate with the bacterial chromosome; such a replication control is called ***stringent control***.

25.5.2.2. Multicopy Plasmids

Each bacterial cell contains more than one copy per cell, the number of copies being characteristic for individual plasmids. Their replication control is such that it allows more than one round of plasmid replication to occur for every replication of host cell chromosome. This type of replication control is often referred to as ***relaxed control***. Plasmid ColE1 is maintained in ~ 20 copies per *E. coli* cell.

25.5.3. Compatibility Groups

A ***plasmid compatibility group*** includes all those plasmids whose members can not be maintained together in the same host cell. Therefore, those plasmids that can be maintained together in the same cell belong to different compatibility groups. It appears that the replication control of all the plasmids belonging to the same compatibility group is identical. As a result, the replication control system regards the members of a single compatibility group as copies of a single plasmid, and the replication of all such plasmids is stopped till the characteristic copy member is restored. Thus plasmids appear to be ***selfish elements***; a plasmid residing in a bacterial cell uses the incompatibility devices to prevent other similar plasmids from occupying the same cell.

25.5.4. Segregation and Maintenance

Plasmids employ several types of mechanisms to ensure their maintenance in bacterial populations; these include (1) site-specific recombination systems, (2) addiction systems (the cell that lose plasmid dies), (3) and true partition systems (segregation similar to chromosomes; found in case of all low copy number plasmids). Any further discussion of these mechanisms is beyond the scope of this book.

Transposable Elements

- DNA sequence that change their position in genome.
- *Types*: (1) Sequences that transpose as DNA (*insertion sequences* and *composite transposes*); and (2) Sequences that transpose as RNA intermediate (*retroposons*).

25.6. TRANSPOSABLE ELEMENTS

Transposable elements are DNA sequences that move from one location in a chromosome to another location within the same chromosome or into another chromosome. In bacteria, these elements can move from bacterial chromosome into a bacteriophage chromosome or into a plasmid and *vice-versa*. Transposable elements are of the followings two main types.

1. One group of transposable elements propagates itself by transposition as DNA copies, *e.g.*, insertion sequences (IS) and composite transposons.
2. Another group of transposable elements propagates itself by making DNA copies of their RNA genomes; the DNA copy is transcribed to yield the RNA genomes, *e.g.*, retroposons (Section 7).

Insertion Sequences

- Sequences of 800–1,400 bp.
- 9 – 24 bp inverse repeats at their ends.
- Single coding region encodes transposase.
- In *E. coli*, occur at many but nonrandom sites.
- Promote illegitimate recombination.
- Insertion within a gene prevents its function.
- Integration generates 5–9 bp repeats in target DNA.

25.6.1. Insertion Sequences (IS)

IS elements are short DNA sequences of about 800 – 1,400 bp and have the ability to ***transpose***, *i.e.*, move from one position to another in the genome. Examples of such sequences are *E. coli* IS elements *IS*1, *IS*2 *IS*3, *IS*4, etc. They are the simplest forms of transposable elements; their name is due to their discovery as insertions within bacterial operons that prevented transcription or translation of the concerned operons/genes.

IS elements have 9 – 24 bp inverse repeats (the same sequence present in the opposite orientations) at their two ends. The repeated sequences at the two ends of an IS element are closely related rather than identical. These elements contain a single long coding region, which codes for the enzyme transposase, which is responsible for transposition of these elements (Fig. 25.6).

IS elements occur at many sites in the *E. coli* chromosome, but their distribution is nonrandom. Insertion of an IS element into a chromosome generates a short (usually, 5 or 9 bp) direct repeat that flanks the IS element. When this IS element transposes, this direct repeat remains at the site of insertion in the host chromosome. Integration of an IS element within a gene prevents the function of this gene, thereby causing mutation. IS elements are believed to be responsible for an appreciable fraction of spontaneous gene mutations in bacteria and bacteriophages. When two nonhomologous DNA segments contain the same IS element, recombination can occur between their IS elements. Thus IS elements promote recombination between nonhomologous chromosomes; this is called ***illegitimate recombination***.

25.6.2. Composite Transposons

Composite transposons (T_n) are usually > 2,000 bp long, have one or more genes in addition to that coding for transposase and have IS elements repeated at

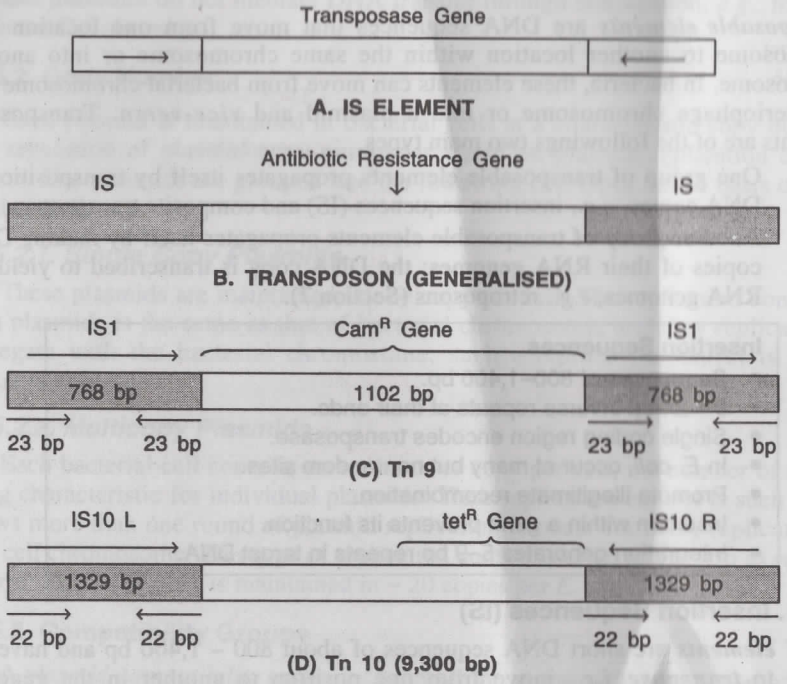


Fig. 25.6. The structure of IS and *Tn* elements. (A) An IS element. (B) The structure of a generalised *Tn* element; *IS* may be an *IS* element or an *IS*-like element. (C) the structure of *Tn*9; the arrows at the top denote the orientation of *IS*1 element, while the four arrows below indicate orientations of the inverse terminal repeats of 23 bp of which 18 bases are the same but 5 are different. (D) Structrue of *Tn*10; arrows at the top denote the orientation of *IS*10, while the four arrows below denote the 22 bp long inverse terminal repeats of which 17 beases are the same, but five are different. (cam^R, chloromphenicol resistance; tet^R, tetracycline resistance).

their ends (Fig. 25.6). The IS elements making up the two ends of a T_n element may be either in the same orientation or in opposite orientations. The IS elements form the two arms of a transposon, while the other genes are present in its central region. Composite transposons are found in bacteria as well as in eukaryotes. Integration of a transposon into a chromosome is accompanied with a duplication of usually 5 – 9 bp of the host chromosome; the repeats are located at the two ends of the T_n element (Fig. 25.7).

The transposition of composite transposons is sponsored by either one or any one of the two IS elements located at their ends. Only the ends of transposons are needed for transposition; in fact, any DNA sequence located between two IS elements can be transposed by these IS elements. The general events during insertion of transposons are follows: (1) staggered breaks are produced in the target DNA, (2) the transposon joins to the protruding single-stranded ends, and (3) the remaining gaps are then filled generating the repeats of target DNA at the site of

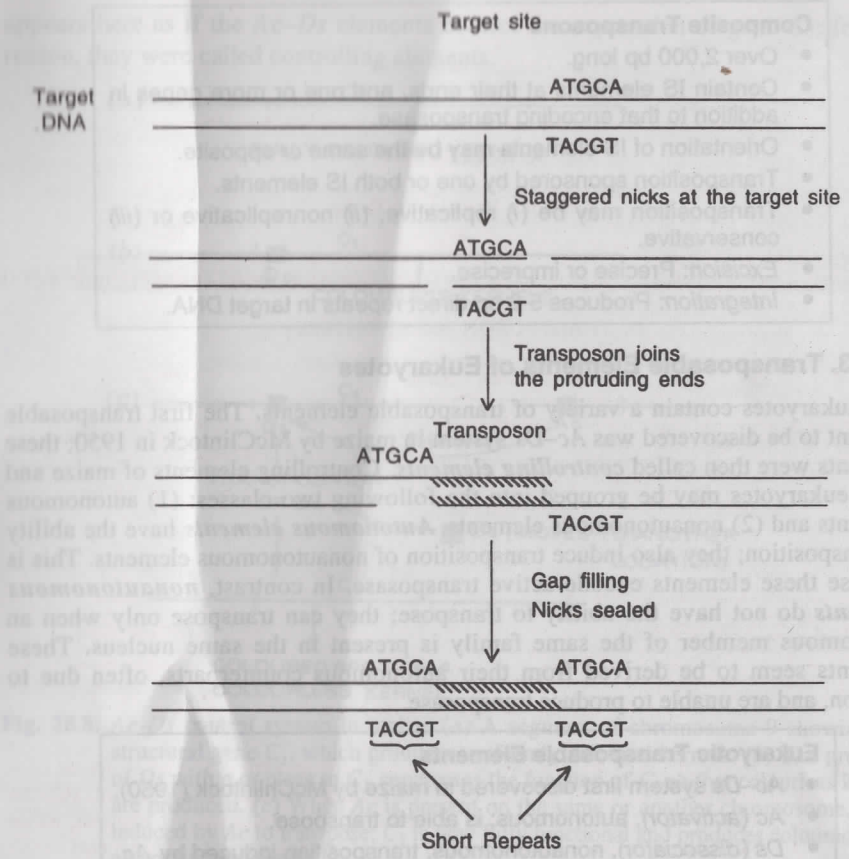


Fig. 25.7. A schematic representation of insertion of transposons giving rise to short repeats of the target site sequence.

insertion (Fig. 25.7). Transposition may be (1) conservative, (2) nonreplicative or (3) replicative.

1. In the case of ***nonreplicative transposition***, the transposon moves out of one site and becomes inserted at a new site. It leaves behind a double–strand break at the old site; the break may or may not be repaired.
2. In the case of ***replicative transposition***, a new copy of the transposon is produced; this copy becomes inserted at the new site. As a result, the original site remains unchanged.
3. In case of ***conservative transposition***, a transposon moves to a new site in such a way that a double–strand break is not produced at the old site.

The excision of transposons may be either precise or imprecise. In ***precise excision***, the transposon along with a copy of the duplicated sequence is removed; this is rare. In ***imprecise excision***, a remanant of the transposon is left at the site; this may prevent reactivation of the gene into which the transposon was located earlier. But precise excision always leads to activation of the affected gene.

Composite Transposons

- Over 2,000 bp long.
- Contain IS elements at their ends, and one or more genes in addition to that encoding transposase.
- Orientation of IS elements may be the same or opposite.
- Transposition sponsored by one or both IS elements.
- Transposition may be (*i*) replicative, (*ii*) nonreplicative or (*iii*) conservative.
- *Excision*: Precise or imprecise.
- *Integration*: Produces 5-9 bp direct repeats in target DNA.

25.6.3. Transposable Elements of Eukaryotes

Eukaryotes contain a variety of transposable elements. The first transposable element to be discovered was *Ac–Ds* system in maize by McClintock in 1950; these elements were then called ***controlling elements***. Controlling elements of maize and other eukaryotes may be grouped into the following two classes: (1) autonomous elements and (2) nonautonomous elements. ***Autonomous elements*** have the ability of transposition; they also induce transposition of nonautonomous elements. This is because these elements encode active transposase. In contrast, ***nonautonomous elements*** do not have the ability to transpose; they can transpose only when an autonomous member of the same family is present in the same nucleus. These elements seem to be derived from their autonomous counterparts, often due to deletion, and are unable to produce transposase.

Eukaryotic Transposable Elements

- *Ac–Ds* system first discovered in maize by McChlintock (1950).
- *Ac* (*activator*), autonomous; is able to transpose.
- *Ds* (*dissociator*), nonautonomous, transposition induced by *Ac*.
- *Ds* derived from *Ac* by deletion.
- Transposition occurs following DNA replication.
- Insertion within a gene prevents its function; transpositibn out of the gne allows function.

In the *Ac-Ds* system, *Ac* (activator) is the autonomous element, while *Ds* (dissociator) is the nonautonomous element. *Ds* is derived from *Ac* by deletion of 194 bp to 2.5 kb. *Ac* has 11 bp repeats at the ends, encodes a single protein, possibly transposase, and transposes by a nonreplicative mechanism. Usually, the new site of insertion is on the same chromosome, close to the old site. Transposition occurs following DNA replication, and the donor site suffers a double–strand break.

In maize, gene C_1 produces coloured kernels. If *Ds* element (or even *Ac*) becomes inserted within the gene C_1, this gene can not function. As a result, colourless kernel are produced. In case *Ac* is present in the same nucleus, it may induce *Ds* to transpose out of the gene C_1 to a new location. This will allow gene C_1 to express and produce colour (Fig. 25.8). Transposition of *Ds* during kernel development will generate spots of colour on colourless kernels. The earlier (during kernel development) the occurrence of transposition; the larger is the size of spot. It

appears here as if the *Ac–Ds* elements control the expression of gene C_1; for this reason, they were called controlling elements.

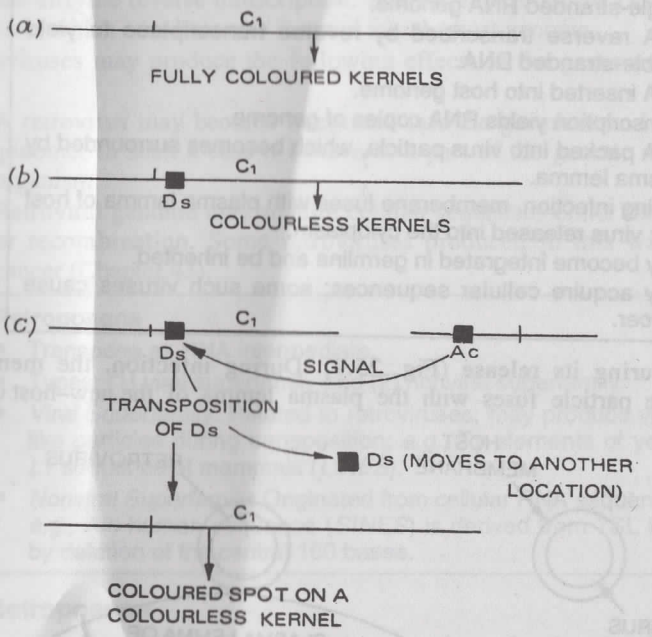


Fig. 25.8. *Ac-Ds* control system in maize. (*a*) A segment of chromosome 9 showing the structural gene C_1, which produces uniformly coloured kernels. (*b*) The presence of *Ds* within or close to C_1 suppresses the function of C_1 so that colourless kernels are produced. (*c*) When *Ac* is present on the same or another chromosome, *Ds* is induced by *Ac* to transpose; C_1 is now fully functional and produces coloured spots on colourless kernels.

Other transposable elements found in maize are *En*, *mu*, *Spm*, etc.; maize contains at least 14 different transposable systems. Another example of eukaryotic transposable element is the *P* element of *Drosophila*.

25.7. RETROVIRUSES AND RETROPOSONS

Retroposons are either DNA copies of RNA viruses, called ***retroviruses*** or derivatives of such DNA copies and may even have lost the ability to transpose. These elements also generate short direct repeats of target DNA during their insertion.

25.7.1. Retroviruses

Retroviruses have single–stranded RNA genomes, which are replicated through a double-stranded DNA intermediate. As a rule, the DNA copy of the RNA genome becomes inserted into the host genome by a transposition like event. The insertion is catalyzed by integrase. The DNA copy of the viral genome is transcribed by the host machinery to yield RNA copies of the viral genome; these copies may function either as mRNA or as viral genome. Two copies of the RNA genome are packaged into each virus particle, which becomes surrounded by a segment of the membrane of

Retroviruses

- Single-stranded RNA genome.
- RNA reverse transcribed by reverse transcriptase to yield double-stranded DNA.
- DNA inserted into host genome.
- Transcription yields RNA copies of genome.
- RNA packed into virus particle, which becomes surrounded by plasma lemma.
- During infection, memberane fuses with plasma lamma of host cell; virus released into the cyloplasm.
- May become integrated in germline and be inherited.
- May acquire cellular sequences; some such viruses cause cancer.

its host cell during its release (Fig. 25.9). During infection, the membrance surrounding the particle fuses with the plasma lemma of the new–host cell; this

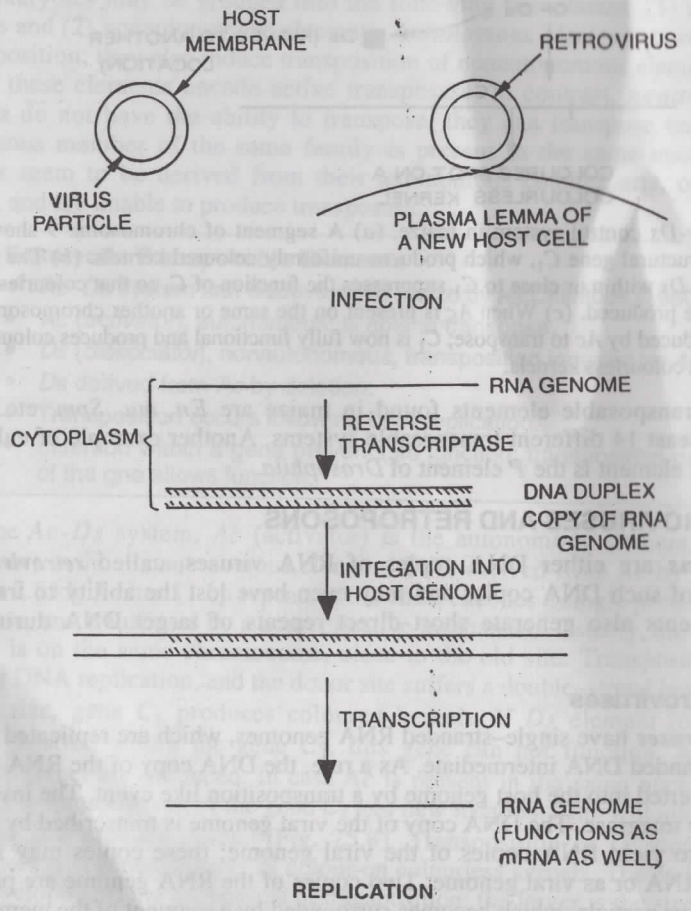


Fig. 25.9. A schematic representation of the infection by and replication of a retrovirus.

releases the virus particle into the host cell cytoplasm. The virus particle carries within it the enzyme reverse transcriptase. This enzyme produces DNA copy of the viral genome, which now becomes inserted into the host genome.

Retroviruses may produce the following effects in the genomes of their host cells.

1. A retrovirus may become integrated into the germline and remain in the genome. In such a case it behave as a part of the genetic material of the organism.
2. Retroviral genome may acquire cellular sequences either through deletion or recombination. Some retroviruses produced in this way may cause cancer (Chapter 43).

Retroposons

- Transpose as RNA intermediate.
- Types: (1) viral superfamily and (2) nonviral superfamily.
- *Viral Superfamily*: Related to retroviruses; may produce virus-like particles during transposition; *e.g.*, Ty elements of yeast, *L*1 sequence of mammals (*LINES*).
- *Nonviral Superfamily*: Originated from cellular RNA sequences; *e.g., Alu* human sequence (*SINES*) is derived from 7SL RNA by deletion of the central 160 bases.

25.7.2. Retroposons

Retroposons are derived from retroviruses and are closely related to them, *e.g.*, Ty element of yeast, *copia and foldback* (*FB*) elements of *D. melanogaster* and *LINES* (long interspersed sequences) and *SINES* (short interspersed sequences) in mammals.

Each Ty element is 6.3 kb long, and has 330 bp direct repeat at its ends (***δ repeats***). Ty elements are of two major types, *viz.*, Ty1 and Ty917. In a typical yeast genome, there are ~30 copies of Ty1 and ~6 copies of Ty917. Ty elements are transcribed into mRNA. These elements have two open reading frames (ORFs) that have a 13 codon long overlap. One ORF encodes a DNA binding protein (TyA), while the other encodes TyB that has homologies with reverse transcriptase, protease and integrase of retroviruses. Ty elements transpose at a frequency of 10^{-7} to 10^{-8} through an RNA intermediate, which is reverse transcribed to yield a DNA duplex that becomes inserted into the yeast genome. They form virus–particles, which accumulate within the cell in which transposition has been induced. Only some of the yeast Ty elements are active, while most of them have lost the ability to transpose.

In mammals, a significant part of moderately repetitive DNA consists of retroposons. Most of these sequences belong to the families called *LINES* and *SINES*. An example of a *LIENS* sequence is L1. This sequence is found in 20 to 50 thousand copies per genome. A typical L1 element is ~6,500 bp long and seems to have originated as a mobile gene coding for its own transposition. L1 is transcribed, has two open reading frames that overlap by 14 bp, and one ORF has a region homologous to reverse transcriptase. *LINES* sequences have originated from RNA polymerase II transcripts. As a result, they lack promoter sequence, and they are often called ***processed pseudogenes***.

An example of ***SINES sequence*** is the *Alu* family of human genome. These sequences are ~300 bp long, occur in ~30,000 copies per human genome and are widely dispersed in the genome. They are called *Alu* sequences because the restriction enzyme *Alu* cleaves them at a single site located at 170 bp. *Alu* sequence is related to 7SL RNA; the central 160 bases (bases 131 to 191) of 7SL RNA are missing from the *Alu* element. *SINES* sequences are derived from transcripts of RNA polymerase III, and may carry internal promoters. The presence of *Alu* sequences in the form of inverted repeats seems to be responsible for most of the secondary structure of mammalian nuclear RNAs. It is possible that *Alu* sequences could be involved in origin of replication function in eukaryotic genomes.

Eukaryotic retroposons may be grouped into the following two categories: (1) ***viral superfamily*** consisting of elements resembling retroviruses, *e.g.*, Ty, L1, etc., and (2) ***nonviral superfamily*** having elements that originated from cellular RNA sequences, *e.g.*, *Alu*, etc.

25.8. PACKAGING OF GENETIC MATERIAL

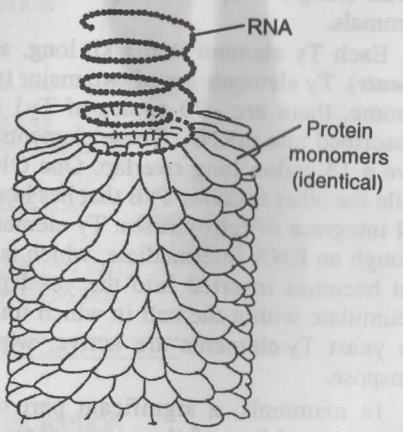
The genetic material is packaged into compact forms in association with basic proteins (DNA and RNA have negative charge); this facilitates its transmission. Packaging condenses DNA and genetic RNA into compact structures typified by (1) virus particles, (2) bacterial nucleoids and (3) eukaryotic chromosomes.

Packaging of Viral Genomes
- Protein capsid assembled around DNA or RNA, *e.g.*, TMV.
- Capsid is first assembled; DNA is then inserted into this capsid; *e.g.*, λ phage.

25.8.1. Packaging of Viral Genomes

Viral genomes are packaged into capsids, which are either like a rod or a spherical-appearing polyhedron, and are assembled from only one or a few types of protein monomers. The packaging of genome follows two general patterns as follows.

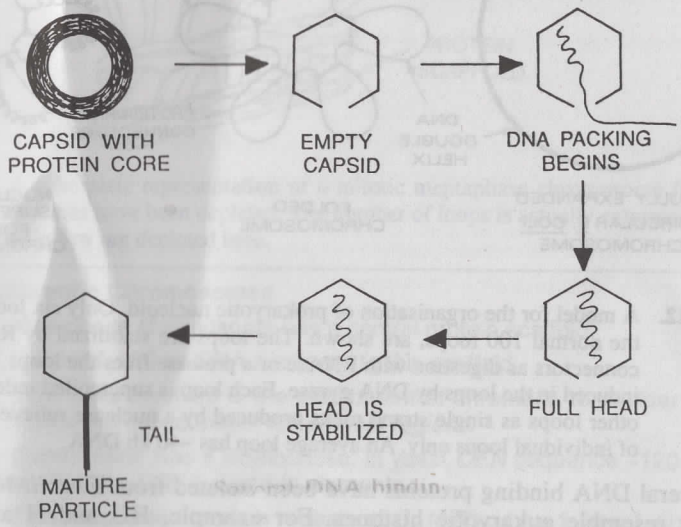
1. *The protein shell is assembled around the DNA or RNA genome by protein-nucleic acid interaction, e.g.,* in TMV. The assembly starts at a duplex hairpin within the RNA genome of TMV; this hairpin is called ***nucleation centre***. From this centre, assembly proceeds in both the directions. The unit of capsid is a two layer disc, each layer containing 17 identical protein subunits. The disc ultimately becomes helical on association with RNA, which forms a coil inside the protein shell so formed (Fig. 25.10). The RNA–protein interaction determines the position of RNA within the capsid.



TMV packaging

Fig. 25.10 Packaging of the single-stranded RNA of TMV genome. Association with the protein units generates the spiral arrangement of the RNA.

2. *The capsid is assembled first and nucleic acid is then inserted into the capsid.* The viral genome becomes condensed as it is inserted into the capsid. In case of DNA viruses, spherical capsids are first assembled from a set of proteins. Subsequently, the capsid becomes empty by losing its protein core and acquires its typical shape (Fig. 25.11). The genome of phage λ occurs as a ***concatemeric molecule*** in which several genomes are joined end-to-end. The ends of genomes are marked by specific sequences called *cos* sites. A cleavage on the left of a *cos* site produces the free end that is first inserted into the capsid. The genome is inserted into the capsid till the next *cos* site located at the right is reached; a cleavage just before the *cos* site generates the other end of the genome. The packaging of DNA into the capsid does not depend on any specific sequence. In fact, any DNA sequence located between two *cos* sites becomes packaged. However, only the usual length of lambda genome is packaged into a capsid.



Fig,. 25.11. A schematic representation of the packaging of DNA duplex in the lambda phage head.

Nucleoids
- In prokaryotes, including blue-green algae.
- *E. coli* chromosome has 100 loops.
- Each loop supercoiled independently.
- Folds stabilized by protein/RNA.
- *E. coli* chromosome packed in 1–2 μ dia nucleoid.
- *E.* coli has histone like proteins, *e.g.*, H, HU.

25.8.2. Nucleoids of Prokaryotes

The *E. coli* chromosome is 1100 μ long; it is coiled and folded to be accommodated in a cell that has diameter of 1-2 μ; such a chromosome is called

folded genome. Folded genome of *E. coli* has ~100 loops that are stabilized by some molecules. Each fold in the chromosome is supercoiled independent of the other (Fig. 25.12). Formation of loops and supercoiling within loops packs the *E. coli* chromosome within a Feulgen positive ***nucleoid*** of 1–2 μ. The nucleoid appears as a mass of transversely and tangentially cut thin fibrils in electron micrographs. The nucleoid is stabilized by the protein and RNA molecules associated with it. Genomes of all prokaryotes, including blue-green algae, are packaged in a similar manner. The nucleoid occupies about 1/3 volume of *E. coli* ceil; it is probably attached to the plasma lamma, which is involved in segregation of the genome during cell division.

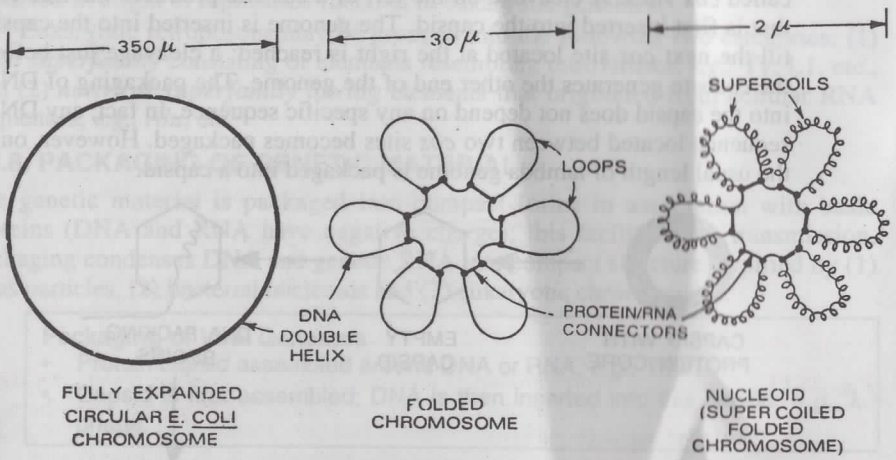


Fig. 25.12. A model for the organisation of prokaryotic nucleoid. Only six loops, in place of the normal 100 loops, are shown. The loops are stabilized by RNA or protein connectors as digestion with RNAse or a protease frees the loops. Supercoils are induced in the loops by DNA gyrase. Each loop is supercoiled independent of the other loops as single strand nicks produced by a nuclease relieve the supercoils of individual loops only. An average loop has ~40 kb DNA.

Several DNA binding proteins have been isolated from *E. coli*. Many of these proteins resemble eukaryotic histones. For example, HU and H are similar to histones H_2b and H_2a, respectively; they are highly basic and bind to DNA. However, the role of these proteins in nucleoid formation is not well known.

25.8.3. Eukaryotic Chromosomes

Genetic material in eukaryotes is organized in chromosomes composed of 300Å diameter chromatin fibers. The chromatin fibers are organized according to nucleosome–solenoid model (Chapter 3, Section 3.11). About 6 μ DNA is packed into 1μ of nucleosome fibre, but 1μ of chromatin fiber contains ~40μ of DNA. But euchromatic regions of metaphase chromosomes contain >1000 μ DNA per micron of their length, while the average for the entire chromosomes is around 10,000 μ DNA per micron of metaphase chromosome.

The organization of metaphase chromosomes seems to be as follows. A central scaffold is formed by ~8% of the proteins contained in the chromosome. Loops of

about 60 kb DNA are anchored in this scaffold (Fig. 25.13). The scaffold looks like a pair of chromatids and consists of a network of fibers; the sister scaffolds are usually tightly connected.

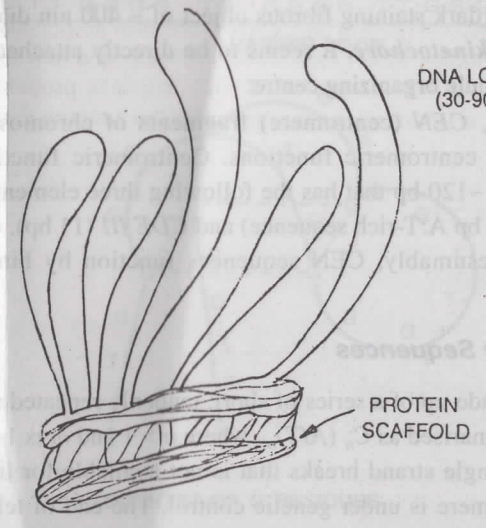


Fig. 25.13. A schematic representation of a mitotic meptaphase chromosome from which histones have been depleted. The number of loops is actually extremely large but only few are depicted here.

Eukaryotic Chromosomes

- Metaphase chromosome has a central protein scaffold.
- Loops of ~ 60 kb DNA anchored to this scaffold.
- DNA sites involved is the attachment of chromatin to nuclear matrix are called *matrix attachment regions*.
- *Centromere*: Has a kinetochose; in yeast *CEN* sequence ~120 bp that has CDE I, II and III elements.
- *Telomere*: Short tandem repeats of C_n $(A/T)_m$ ($n > 1$, $m = 1$-4); length increased by telomerase; end is single -stranded and folded in a hairpin or a complex structure.

Interphase nuclei contain a filamentous structure on the inside of their envelopes; this is called ***nuclear matrix.*** Chromatin often appears attached to the matrix; the DNA sites involved in the attachment are called ***matrix attachment regions*** or ***scaffold attachment regions.*** It has been suggested that this attachment is necessary for transcription or replication.

During replication, the histone octamers associated with DNA are displaced to allow the DNA to replicate. The displaced octamers are not conserved; they either dissociate completely or, at least, partially. Nucleosomes are formed on the newly replicated DNA as soon as it is produced. The location of the first nucleosome on a DNA molecule in some way sets the spacing for the subsequent nucleosomes.

25.8.3.1. Centromeric Sequences

The chromosome regions on either side of a centromere are rich in sat-DNA, and contain a considerable amount of constitutive hetrochromatin. Within the centromeric region, a dark staining fibrous object of ~ 400 nm diameter is present; this particle is called ***kinetochore***. It seems to be directly attached to microtubule, and serves as microtubule organizing centre.

In case of yeast, *CEN* (centromere) fragments of chromosomes (they lack kineto-chore) exhibit centromeric functions. Centromeric function seems to be located in a stretch of ~120 bp that has the following three elements: *CDE I* (9 bp), central *CDE II* (80–90 bp A:T-rich sequence) and *CDE III* (11 bp). *CDE I* and *III* are highly conserved. Presumably, CEN sequences function by binding to specific centromeric proteins.

25.8.3.2. Telomeric Sequences

Telomeres are made up of a series of short, tandenly repeated sequences (Table 25.2) that can be summarised as C_n $(A/T)_m$, where $n > 1$ and m is 1–4. The telomeric region has a set of single strand breaks that is not available for ligase action. The overall length of telomere is under genetic control. The end of telomeric region is produced by folding of one strand of the sequence forming either a hair pin loop or a more complex structure (Fig. 25.14); this makes the telomere end resistant to exonuclease action. Telomeric sequences are extended by the enzyme telomerase. The enzyme telomerase is a ribonucleprotein that has the following two unusual features.

1. Its RNA component has the sequence CCCCAA (in case *Tetrahymena*), which serves as template for the synthesis of TTGGGG repeats added to the telomeric ends.
2. It is a special kind of reverse transcriptase that reverse transcribes its own RNA sequence.

TABLE 25.2

The tandem repeat sequences consisting the telomeres of eukaryotic chromosomes. The sequences are represented as going inward from the telomeric end

Organism	*Repeating unit* (5′ – 3′)
Tetrahymena, Paramecium	*CCCCAA*
Stylonchia, Oxytricha	*CCCCAAAA*
Trypanosoma	*CCCTA*
Saccharomyces (yeast)	$C_{2-3}A(CA)_{1-3}$
Arbidopsis	*CCCTAAA*
Homo sapiens (man)	*CCCTAA*

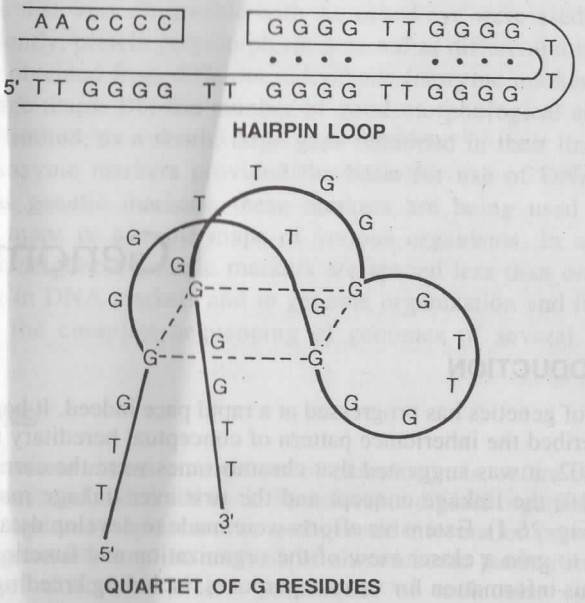


Fig. 25.14. The single-stranded region (G-T rich strand) is folded to form a hairpin, which involves interaction between G base pairs. Alternatively, the second G from 4 repeating units may form a quartet of G residues.

QUESTIONS

1. Differentiate between plasmids and episomes. Give a description of the different types of plasmids and explain the concept and the reasons for plasmid compatibility groups.
2. Give a description of transposable elements, their types, structure and transposition events.
3. Describe briefly the life cycle of retroviruses. Discuss the organization of retroposons and the effects of etroviruses and retroposons on the genome.
4. Briefly describe the various types of repetitive DNA and discuss their distribution and function in the eukaryotic genome.
5. Briefly explain the various organizations of genes in the biological world.
6. Briefly describe the packaging of genetic material in viruses, bacteria and eukaryotes.
7. Write short notes on:
   (*i*) centromeric sequences, (*ii*) telomeric sequences,
   (*iii*) insertion sequences, (*iv*) composite transposable elements,
   (*v*) retroposons, *vi*) retroviruses,
   (*vii*) pseudogenes, (*viii*) split genes,
   (*ix*) overlapping genes, (*x*) gene segments,
   (*xi*) alternative genes, (*xii*) plasmids,
   (*xiii*) episomes, (*xiv*) gene families,
   (*xv*) nucleoids, (*xvi*) repetitive DNA,
   (*xvii*) highly repetitive DNA, (*xviii*) moderately repetitive DNA.

26

Genome Maps

26.1. INTRODUCTION

The science of genetics has progressed at a rapid pace indeed. It began in 1866 when Mendel described the inheritance pattern of conceptual hereditary units, now, called genes. In 1902, it was suggested that chromosomes were the carriers of genes, and soon (in 1910) the linkage concept and the first ever linkage map (in 1913) were developed (Fig. 26.1). Extensive efforts were made to develop detailed linkage maps with a view to gain a closer view of the organization and function of the genomes and to use this information for various purposes, including breeding.

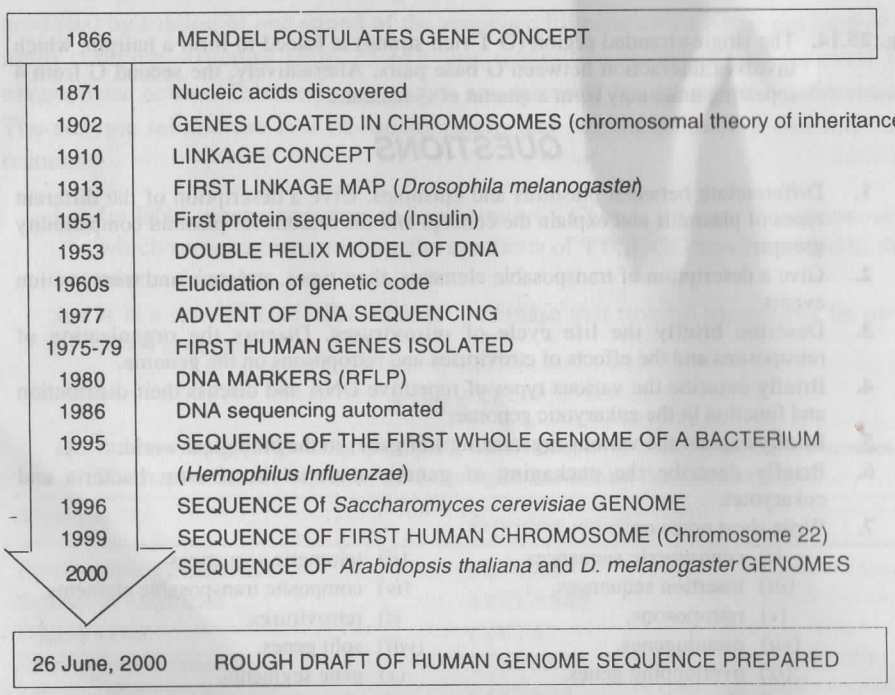


Fig. 26.1. Sequence of events from discovery of genes (hypothetical units of inheritance) by Mendel to the preparation of rough or working draft of human genome sequence. Draft sequence data are mostly in the form of 10,000 bp fragments whose approximate chromosomal locations are known; the base pair sequence of each fragment has been determined at least 4 times.

Initially, traits that were observable with the naked eye were used for linkage mapping. Subsequently, protein polymorphism detected as differential mobilities of the same proteins obtained from different individuals (isozyme markers) was also ordered into linkage maps. But the number of good morphological and isozyme markers is rather limited; as a result, large gaps remained in their linkage maps. The work with isozyme markers provided the basis for use of DNA sequence polymorphisms as genetic markers; these markers are being used to prepare saturated genetic maps or genome maps of various organisms. In a ***saturated genetic map***, next neighbour genetic markers are spaced less than one map unit apart. The interest in DNA markers and in genome organization and function has ultimately led to the complete sequencing of genomes of several organisms, including man.

26.2. DEFINITIONS

In cytogenetics, ***genome*** means a single set of chromosomes, refers ordinarily to nuclear chromosomes, and is denoted by *x*. Therefore, in diploid organisms, ***genome*** (sometimes called ***haplome***) represents the total genetic information present in them. ***Genome analysis*** generally implies analysis of chromosome pairing relationships among genomes of different species; techniques of molecular cytogenitics, cytochemistry and molecular biology have also been employed to determine genomic affinities. But in the modern context, ***genome analysis*** refers to a detailed analysis of the structural and functional organization of the complete genome; this is commonly termed as ***genomics***. In the initial phase, genomics consists of development of genome maps. A ***genome map*** may be defined as a detailed schematic description of the structural and functional organization of the complete genome of an organism. *The ultimate aim of genomics is to obtain the DNA sequence of the complete genome, which provides the most detailed molecular description, i.e., the complete nucleotide sequence, of the genome.*

Genetic marker is a specific location on a chromosome that is defined by a naked eye polymorphism (NEP), or protein or DNA sequence; the inheritance of a genetic marker can be followed in a mapping population. ***Naked eye polymorphism*** consists of those genetic markers whose inheritance can be followed with the naked eye, and without any specialized biochemical or molecular technique. ***Mapping population*** is any population that is studied for the preparation of genetic maps. An ***isozyme*** is a variant of an enzyme or protein that shows a different mobility during electrophoresis, and can be used as a genetic marker. ***Random amplified polymorphic DNA (RAPD)*** is a DNA marker that is defined by differences between individuals in terms of DNA regions either being or not being amplified in a polymerase chain reaction primed by random oligonucleotide (usually, 10-mer) sequences (Section 26.6.3). Another DNA marker, is defined by differences between individuals in terms of the size of fragments generated by restriction enzyme digestion of genomic DNA and detected by hybridization with specific cloned DNA sequences called probes (Section 26.6.2) is termed as ***restriction fragment length polymorphism (RFLP)***. A ***molecular marker*** is a genetic marker that is detected as differential mobility of a protein/DNA fragment or as differential amplification of a DNA sequence present in different individuals.

26.3. GENETIC MAPS

A ***genetic map*** is a schematic representation of the various genetic markers in the specific order in which they are located in a chromosome as well as the relative distances between these markers. Three different strategies have been used to construct genetic maps, which are usually referred to as linkage maps, cytogenetic or cytological maps and physical maps (Fig. 26.2).

26.3.1. Linkage Maps

A ***linkage map*** depicts the order of genetic markers and the relative distances between them as measured in terms of recombination frequencies between the markers. A ***map unit*** or ***centrimorgan (cM)*** is that distance, which allows 1% recombination between the concerned genes. For example, genes *sc* (scute, an abnormality in bristle formation) and *w* (white eye) of *D. melanogaster* are located 1.5 map units apart in the linkage map (Fig. 26.2).

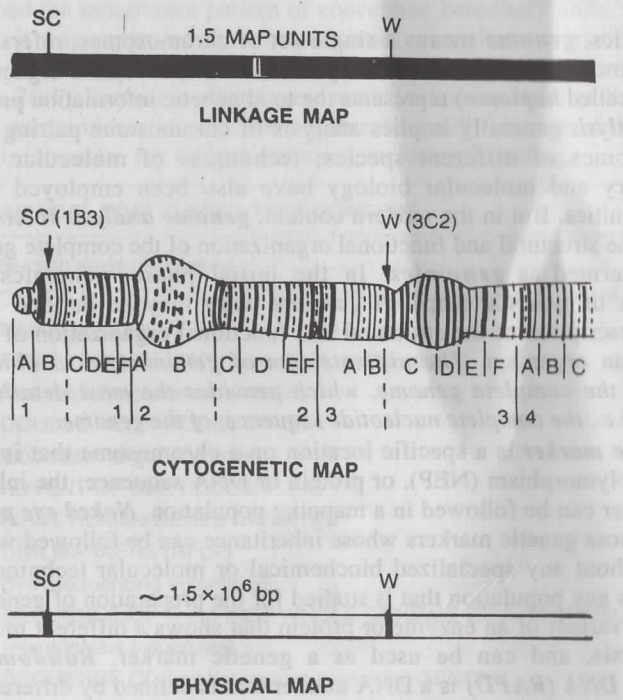


Fig. 26.2. Linkage, cytogeneitic and physical maps of genes *sc* (scute) and *w* (white eye) of *Drosophila melanogaster.*

26.3.2. Cytogenetic Maps

A ***cytogenetic map*** depicts the locations of various genes in a chromosome relative to specific microscopically visible landmarks in the chromosome. In most cases, each chromosome has a characteristic banding pattern, which may be either naturally present, *e.g.*, in polytene chromosomes of *Drosophila,* or more commonly generated by specific staining protocols, *e.g.*, in case of human chromosomes; the genes are mapped cytologically relative to these band locations (Fig. 26.2). Cytogenetic mapping is far more refined in organisms like *Drosophila* because their polytene chromosomes have a very detailed banding pattern. It is commonly used in eukaryotes since they have relatively large microscopically observable chromosomes.

Cytogenetic mapping is often used as a first step in the localization of genes in plants and animals, which is then followed by physical mapping. In most species, cytogenetic mapping is only accurate within limits of ~5 megabase pairs (Mb = 10^6 bp) along a chromosome; the resolution is much better in species that have polytene chromosomes. For example, the *sc* gene of *Drosophila* is located at band *1B3* of X chromosome, while gene *w* is located in band *3c2*. Cytogenetic mapping can be done using any one or more of several approaches, such as fluorescent *in situ* hybridization (FISH), somatic cell hybridization and analysis of changes in polytene chromosomes.

26.3.3. Physical Maps

In a ***physical map,*** genes are depicted in the same order as they occur in chromosome, and the distances between them are shown as number of base pairs separating the genes. For example, genes *sc* and *w* of *Drosophila* are separated by ~1.5×10^6 bp (Fig. 26.1). In contrast, the distances between genes in a linkage map are shown as the frequency of recombination between them. Physical mapping usually requires the (1) cloning of many pieces of chromosomal DNA, (2) characterization of these fragments for size, *i.e.*, length in base pairs, and the genes present in them, and (3) determination of their relative location along a chromosome, (*e.g.*, by McFISH). *The ultimate physical mapping is the complete sequencing of a genome.*

Correlations between linkage, cytogenetic and physical maps often vary from one species to another and from one region of a chromosome to another region of the same chromosome. For example, 1 cM (= map unit) represents 1,000 kb of DNA in humans, while in mice it is equivalent to ~1,600 kb of DNA. In case of maize, 1 cM equals an average physical distance of ~1,500 kb, while in *Arabidopsis* it has been estimated as 200 kb only. Similarly, 1 cM may represent only 43 kb in one region of a tomato chromosome, but in the other regions it may be equivalent to a much longer (upto 600 kb) segment of DNA. These differences arise due to the differences in the frequencies of recombination in different species and/or chromosome regions. In general, recombination is inhibited near centromeres, in heterochromatin and in regions introgressed from related species. Therefore, linkage maps have many DNA markers clustered near centromeres and in heterochromatic regions.

26.4. GENETIC MARKERS

Any trait of an organism that can be identified with confidence and relative ease, and can be followed in a mapping population is called ***genetic marker.*** Genetic markers can be detected with the naked eye (naked eye polymorphism), as differences in electrophoretic mobilities of specific proteins, or as differences in specific DNA sequences (detectable as RFLP or differential amplification by PCR). A comparison of the salient features of different markers is given in Table 26.1.

TABLE 26.1
A comparison among the salient features of NEP (naked eye polymorphism), protein-based and DNA markers

Feature	*NEP markers*	*Isozyme markers*	*DNA markers*
Efforts required for detection	Simple (visual observation)	Moderate	Moderate to difficult
Ease in use	Easy	Moderately difficult	Moderately difficult to difficult
Number of different markers per cross	< 10	< 30	Unlimited
Plant material required for detection	Intact plant	Little amount of tissue	Little to medium amount of tissue
Reproducibility	High	High	Moderate to high
Dominant/codominant	Generally, dominant	Codominant	Dominant (RAPD), codominant
Feature of the organism scored	Phenotype	Protein product	DNA base sequence
Biological meaning of the markers	Consequences of gene action	Genes that are expressed	Arbitrary DNA sequences, may or may not represent genes

26.4.1. Naked Eye Polymorphisms (NEP)

Those genetic markers whose inheritance can be followed with the naked eye constitute ***nake-eye polymorphisms.*** The traits included in this group are plant height, disease response, photoperiod sensitivity, shape or colour of flowers, fruits or seeds and many other visually scorable traits. (1) NEP traits represent the actual phenotypes of importance to us, while protein and DNA markers are important only as arbitrary loci used for linkage mapping and often do not correspond directly to specific phenotypes. Further, (2) NEPs can be generally scored quickly, simply and without laboratory equipment. Typically, (3) only a few NEP factors can be analyzed in a single population; the number of segregating factors is constrained mainly by the difficulty in determining the phenotype of several NEPs in a single plant. In early days, two or three-factor crosses were widely used, and data from several crosses were pooled to construct linkage groups. As long as two or more loci were common between different mapping populations, *i.e.,* crosses, the gene order and distances could to integrated with some degree of confidence.

26.4.2. Protein-Based Markers

These markers are detected as electrophoretic variants of proteins. Usually, variants of enzymes, called ***isozymes,*** are used as markers. But variants in the molecular weight or isoelectric point of structural proteins, especially seed storage proteins, can also be used as genetic markers. Protein markers offer the following advantages over NEP markers: (1) they reflect differences in gene sequence more directly than do NEPs, (2) only a small amount of tissue is needed for their detection, (3) they can often be detected when the plant is very young or even directly from seeds, and (4) analysis of one marker usually does not interfere with the ability to assay other protein markers. Therefore, many different protein loci can be followed in the same cross. However, (5) only a relatively small number of

protein polymorphisms may exist between two parents; this limits the total number of protein markers that can be actually scored in a given cross.

26.4.3. DNA Markers

In 1980, scientists studying human genetics observed that variation in the pattern of DNA fragments generated by restriction enzyme digestion of genomic DNA could be used as genetic marker. Soon this approach was extended to plants; at present, DNA markers represent the primary method of genetic mapping. There are several types of DNA markers, which can be grouped into the following two broad categories: (1) restriction fragment length polymorphism (RFLP), and (2) PCR-based DNA markers; some DNA markers, *e.g.,* amplified fragment length polymorphism, combine some features of both the above marker types.

DNA markers (1) represent polymorphism in the actual base sequence of DNA. (2) Scoring for one DNA marker usually has no effect on that of the others, and essentially there is no limit to the number of DNA markers that are available for scoring in a single mapping population. As a result, highly saturated DNA marker maps have been developed for several crop species, *e.g.,* rice, maize, common bean, lettuce, potato, soybean, tomato, wheat, etc. (3) But such factors as amount of DNA available and the overall level of DNA sequence polymorphism present between the two parents affect the number of different DNA markers that can reasonably be scored in a single mapping population.

26.5. THE HUMAN GENOME PROJECT

This is an undertaking by many countries (currently administered jointly by the National Institute of Health and Department of Energy, U.S.A.) to acquire "complete knowledge of the organisation, structure, and function of the human genome". It is called *International Human Genome Sequencing Consortium* and is aimed at finding out all the genes in each of the human chromosomes (and ultimately of other organisms), determining their function, and hopefully understanding how they together form the complete organism. This is regarded as the most ambitious project ever undertaken by man. Discussions among scientists to undertake this project began in the mid 1980s. In 1988, the National Institute of Health, Bethesda, Maryland established an Office of Human Genome Project with James Watson as its first director. The Human Genome Project officially began on October 1, 1990. The goals of the genome project are as follows.

1. *To develop a genetic linkage map of human genome* by identifying thousands of genetic markers and mapping them in the genome. As of August 2000, over 9,300 markers had been mapped to particular chromosomes.
2. *To obtain a physical map of human genome* by cloning genomic DNA into YACs and cosmids.
3. *To sequence the entire human genome.* The euchromatic region of human genome has ~3.2 billion base pairs; it would take more than 1 million pages of this book to type this sequence. As on Feb. 11, 2001, about 93.5% of human genome had been sequenced of which only 32.5% is finished, while the remaining 61.0% is as draft. The ***finished sequence*** is

in form of contigs, and the base sequence has been determined 8 to 9 times. But in ***draft sequence,*** the sequences are mostly in the form of 10,000 bp fragments whose approximate chromosomal locations are known, and gaps exist between fragments. Complete sequence is known for only chromosomes 21 and 22. The complete sequence of human genome euchromatin is expected to become available by 2003.

4. *To develop technology for the management of human genome information,* by developing user-friendly tools to provide scientists easy access to up-to-date information generated by the project, and by developing analytical tools to interpret this information.
5. *To analyze the genome of other model organisms,* e.g., E. coli, D. melanogaster, Arabidopsis thaliana, Mus musculus (*mouse*), *etc.*
6. *To develop programmes focussed on understanding and addressing the ethical, legal, and social implications of the results obtained from the project.*
7. *To develop technological advances in genetic methodologies, e.g.,* gene cloning, contig construction, DNA sequencing, etc. These advances have in fact occurred so that the complete sequence of human genome was obtained in 2001, much before the projected time.

A great benefit expected from the project is the ability to identify human genes. The potential identification of genes that are mutated in the development of disease is a strong motivation for this project. Further, the complete genome sequence will enable researchers an insight into the types of proteins encoded by these genes. The cloning and sequencing of the disease causing alleles is expected to greatly facilitate diagnosis and treatment of diseases.

The basic strategy of the genome project may be grouped into the following three stages: (1) mapping, (2) sequencing and (3) functional analysis. It may be pointed out that objectives 1, 2, 3 and 7 of the project have already been achieved.

26.5.1. Mapping

The first major goal of the project is to prepare high resolution or saturated genetic and physical maps of human genome (and genomes of other experimental organisms). Molecular markers have been used to produce maps of all the human chromosomes in which markers are located only a few centimorgans (cM) apart. Therefore, any new DNA sequence (*i.e.,* an additional marker) can be easily linked with these markers. The locations in the map of markers like simple sequence repeats (SSRs) are called ***sequence tagged sites*** (STS) because such sites can be described by a small amount of DNA sequence information, and the marker for which can be easily synthesized. Microsatellites have served as very useful genetic markers in human genome mapping. Linkage mapping in humans is usually based on analysis of very large pedigrees, which allows the determination of recombination frequencies. In addition, few polymorphic genes are available in humans that can be used for mapping. PCR amplification of particular microsatellites provides a powerful genetic marker for human genomes.

26.5.2. Sequencing

It is envisaged to sequence the entire genomic DNA of humans and mice. At present, this is limited because of limitations of the sequencing technique, *e.g.,* duplicative sequencing and organising the sequences obtained into the correct order. The DNA fragments used for sequencing are generally obtained by the "***shot gun approach***", in which DNA fragments generated from a given source, *e.g.,* whole genomic DNA, a YAC, or a cosmid clone, are cloned and sequenced at random. As a result, a DNA fragment is ordinarily sequenced more than once, and the various fragments sequenced belong to random locations in the DNA source. Therefore, integrating the sequence data into the correct order to obtain the sequence of the entire DNA molecule is a prohibitive task even with the enormous computer capabilities. Automated DNA sequencing and more refined techniques have enabled the preparation of draft sequence of human genome in 2000.

New technologies are being sought and developed leading to more complex and sophisticated approaches. The description presented here is, of necessity, extremely simplified and limited to give the students only a general idea of the objectives and the approaches.

26.5.3. Functional Analysis

The ultimate objective of the genome project is to decipher the function of each of the 30,000-40,000 genes estimated to be present in the human genome. This is the most challenging task and may take several decades. There are 3 general approaches for the functional analysis: (1) analysis of cDNA clones from specific tissues, *e.g.,* brain; each cDNA clone is sequenced only for a short region, which provides an STS for the concerned gene as well as enough base sequence data to decide if the gene is new or already known. This way, it should be possible to determine the tissue-specific and ubiquitous (functional in all tissues) genes, and also to map the genes. (2) The sequence data may be analysed using a computer to identify putative genes present in the genome. Finally, (3) The gene function has to be determined by studying the effects of mutation. The present techniques for site-directed mutagenesis are labour-intensive, and new more convenient methods need to be devised.

26.6. DNA MARKERS

Isozymes (electrophoretic variants of enzymes) were and DNA sequences are used as molecular markers in chromosome mapping. Therefore, for all practical purposes, a ***molecular marker*** may be defined as a DNA sequence used for chromosome mapping as it can be located at a specific site in a chromosome. A molecular marker may be either (1) anonymous or (2) defined. An ***anonymous marker*** is a cloned random DNA fragment whose function or specific features are not known. But a ***defined marker*** may contain a gene, or some other specific feature, *e.g.,* restriction sites for rare cutting restriction enzymes, etc. Such markers are said to be ***polymorphic markers*** if they have different structures in different individuals of a species; the structural differences in a given marker are usually detectable as differential mobilities in gel electrophoresis. *As a practical rule of thumb, a DNA segment is considered monomorphic when over 99% of the individuals in a population have identical sequences at that segment.* The polymorphic molecular markers are of several types of which the following are considered in some detail:

(1) restriction fragment length polymorphism (RFLP), (2) random amplified polymorphic (DNAs (RAPDs), (3) amplified fragment length polymorphism (AFLP) and (4) sequence–tagged sites (STSs).

26.6.1. Restriction Fragment Length Polymorphism (RFLP)

RFLP is generated by the presence and absence of a recognition site for the same restriction endonuclease in the same region of a chromosome from the different individuals of a species. As a result, the concerned restriction enzyme produces fragments of different lengths representing the same chromosome region of different individuals; these differences are detected by gel electrophoresis combined with hybridization with a labelled (usually, radiolabelled) probe specific for that chromosome region. Thus RFLP markers result from a combination of a specific restriction endonuclease and a specific DNA sequence used as probe; a more detailed description of RFLP is given in Section 26.7.

26.6.2. Random Amplified Polymorphic DNAs (RAPDs)

RAPD polymorphism is detected by using oligonucleotides, usually 10 bases long, of random sequences as primers in a PCR reaction. In a strain that has in its genomic DNA sequences complementary to the primer oligonucleotide, PCR products will be detected in the gel, while in those strains that do not have the complementary sequence, no product will be detected. In practice, some bands do appear in each of the strains tested because the stringency of pairing between primers and the template DNA is reduced to ensure amplification. Some of these bands are present in some and absent in the other strains; these bands constitute the RAPD loci. Thus RAPD is generated by specific oligonucleotides used in a PCR reaction. The RAPDs may be of different types, *viz.*, AP-PCR (Arbitrary Primer-PCR), AMP-RLP these are PCR amplified fragment lengths consisting of VNTRs), DAF (DNA Amplification Fingerprinting), MVR-PCR (Minisatellite Variant Repeat-PCR), etc. These procedures differ in the length of primers used, the amplification conditions, and the resolution and the visualization of PCR products. RAPD is described in some detail in Chapter 27. RAPDs are more sensitive than RFLPs to experimental conditions making them more difficult to be consistent and reproducible.

Typical RFLP and RAPD markers show limited variation between parents, especially in naturally inbreeding species. This limits the number of these markers that can be effectively mapped in a single cross. In contrast, micro- and mini-satellites (also called ***simple sequence repeats, SSRs,*** and ***variable number of tandem repeats, VNTRs***) tend to be hypervariable; these sequences are used in mapping by converting them into sequence-tagged sites.

26.6.3. Amplified Fragment Length Polymorphism (AFLP)

AFLP shares some features of both RFLP and RAPD analyses. It uses restriction enzyme-digested genomic DNA as template for PCR amplification using primers that contain the restriction enzyme recognition sites plus a number of, usually 2-3, arbitrary nucleotides. The amplified products are then resolved by denaturing polyacrylamide gel electrophoresis (PAGE). This approach allows the simultaneous screening of a large number of anonymous markers randomly distributed throughout the genome. AFLPs are faster, less labour intensive and provide more information than RFLPs, and they are highly reproducible, which is a

great advantage over RAPDs. In fact, some workers consider the great quantities of information generated by AFLPs as its major drawback. An ***AFLP*** is detected as a differential mobility of an amplified DNA fragment; AFLPs, therefore, behave as codominant markers during inheritance.

TABLE 26.3
A comparison among some of the different DNA markers used in genetic mapping

Feature	*DNA Marker*			
	RFLP	*RAPD*	*AFLP*	*STS*
Efforts to generate marker	Moderate	Simple	Simpler than RFLPs	Difficult
Ease in use	Difficult	Easy	Moderate	Moderate
Number of markers scorable per cross	Unlimited	Unlimited	Unlimited	Unlimited
Plant material required	Medium	Little	Medium	Little
Reproducibility	High	Moderate	High	High
Use of radioactivity	Yes	No	Yes/No	No
Use of restriction enzymes	Yes	No	Yes	No
Use of PCR	No	Yes	Yes	Yes
Dominance	Codominant	Dominant	Codominant	Codominant/ Dominant*
Information generated	Moderate to high	High	Very high	Very high
Time required for detection	Relatively longer	Short	Short	Short

* Codominant in the case of polymorphic STSs and STSs developed from RFLP clones. Dominant for unique STSs and STSs created from RAPD amplification products.

26.6.4. Sequence-Tagged Sites (STSs)

A ***sequence-tagged site (STS)*** is any site on the genome that is unambiguously defined in terms of flanking primers that are used for PCR amplification of this site. Thus when a pair of PCR primers amplifies only a single sequence within a haploid genome, the amplified region of the genome is called a ***sequence tagged site***. STSs can be created in the following ways.

1. Sequencing the two ends of a PCR amplification product (***amplicon*** = the DNA sequence amplified using a set of primers) beyond the ends of the RAPD primer makes it possible to create improved primer sequences that are far more reproducible than are RAPDs.
2. The ends of an RFLP DNA fragment may be sequenced and the information can be used to create primers that allow PCR amplification of this RFLP locus, and thus convert this RFLP into an STS locus.
3. STSs are often created by determining the unique sequences flanking mini- and microsatellite sites and using this information to design primers for PCR amplification of these sites.
4. Sequences of 200-400 bp long DNA segments obtained randomly from the genome may be determined and primers may be designed to amplify these sites in the genome. A proportion of these sites may turn out to be useful STSs.

Thus the creation of STS markers requires considerable amount of work but their use in mapping is dependent on only the knowledge of their primer sequences. STSs based on mini- and micro-satellite sequences and those created from random DNA segments are considered here briefly.

26.6.4.1. Polymorphic Sequence-Tagged Sites

These STSs are based on DNA sequences that are called ***variable number of tandem repeats (VNTRs)***. VNTR sequences are dispersed throughout the genome and are made up of a variable number of end-to-end duplications of identical or almost identical sequences of 2-80 bp each. In this consideration, the VNTR present at a given site in a chromosome is considered as a ***locus*** distinct from those present at other sites of the same chromosome or in other chromosomes. *Polymorphism in VNTRs is usually associated with the number of repeats,* i.e., *different individuals show different numbers of repeat units at a given locus or position in a chromosome; these variations constitute the* ***alleles*** *of VNTR loci.* Polymorphism also arises due to a variation in the distribution pattern of repeated sequences in the different alleles, especially in case of minisatellite DNAs; this is analysed using the MVR-PCR technique. The VNTRs are generally classified into the following two groups: (*i*) minisatellite and (*ii*) microsatellite DNAs; these together constitute the ***hypervariable region*** or ***DNA.***

1. ***Minisatellite DNAs.*** These are usually 0.2 to 2 kb long and consist of tandem repeats of 15 to 60 bp sequences. In case of humans, minisatellite DNAs are concentrated in the proterminal regions of chromosomes. Therefore, they do not constitute a good marker system for human genome despite the great degree of polymorphism exhibited by them.

2. ***Microsatellite DNAs.*** These are short DNA fragments of usually less than 100 bp and are made up of tandem repeats of 2 to 7 bp; therefore, they are also called ***short tandem repeats (STRs), simple sequence repeats (SSRs)*** and ***simple sequence length polymorphism (SSLPs)***. STRs are tandem repeats and are frequently distributed in all eukaryotic genomes (except yeast) examined so far. They show a large, stable polymorphism due to variation in the number of repeat units and are almost ideal as molecular markers for genome mapping. The most common microsatellite sequence encountered in human genome is the sequence (CA)*n*, where *n* may range from 5 to 50 or more. This microsatellite is found, on an average, about every 10 kbp in the human genome. The creation of STSs based on microsatellites is achieved as follows.

In most cases, the (CA)*n* microsatellite is flanked by unique DNA sequences; these sequences are used to create primers for PCR amplification of the microsatellite sites. As a result, a specific pair of primers will amplify only that particular microsatellite site whose flanking sequences were used to create the primers. Since the microsatellites at a single site differ in different individuals in terms of the number of repeats present, the PCR products for a microsatellite STS obtained from different individuals will differ in the mobility of bands. Therefore, STSs show codominance. The PCR products are analysed by high-resolution gel electrophoresis, which can detect small differences in the lengths of DNA fragments. These STSs are very useful in genome mapping and in DNA fingerprinting.

26.6.4.2. Unique Sequence-Tagged Sites

These STSs do not contain microsatellites, behave as dominant markers, and are created as follows. Identification of an STS begins by sequencing a relatively short region, usually ~400 bp, of chromosomal DNA from a genomic library clone. The sequence within this region is used to create a pair of primers that are located 200 to 400 bp apart. The primers, usually 20 bp long, are created by using a computer programme that scans the DNA sequence and identifies regions with optimal primer sequences for PCR. This pair of primers is then used for PCR amplification that contains the total genomic DNA as template. If the PCR produces a single product of the correct size, a unique STS has been identified. In Human Genome Project, about 50% of the primers created in this way identify unique STSs. The unique STSs are useful in creation of contigs required in physical mapping of genomes.

26.7. LINKAGE MAPPING OF RFLP MARKERS

Linkage maps are prepared using recombination frequencies between genes as a measure of distance between them (Chapter 19). We can map RFLPs and other molecular markers much in the same way as we map other genetic markers, *i.e.,* by making crosses between two parents differing for two or more RFLPs and scoring the progeny for these RFLP markers. Let us suppose that one RFLP is detected with probe # 1 and gives bands of 5 kb and 6 kb in different strains. Similarly, the second RFLP is detected by probe # 2 and yields bands of 2 kb and 3 kb. The mapping of these two RFLP markers may be done as follow.

1. Two strains that are homozygous for the two RFLP markers, *e.g.,* one shows bands of 5 kb and 3 kb, while the other shows 6 and 2 kb bands (Fig. 26.4), are crossed together.
2. The F_1 so obtained shows all the four bands, *viz.*, 6, 5, 3 and 2 kb bands. The F_1 is back-crossed to one of the parents, *e.g.,* the parent having 6 and 2 kb bands.
3. The resulting back-cross progeny are scored for these RFLP bands. Each progeny will yield one of the four RFLP band patterns listed in Table 26.4. If the two RFLP markers were not linked, equal proportion of progeny will be present in these 4 categories. But in case of linkage, the proportion of parental progeny will be much higher than that of recombinant progeny.

TABLE 26.4

The pattern of RFLP markers present in back-cross progeny produced by crossing F_1 (showing bands of 6, 5, 3 and 2 kb) with the parent yielding 6 and 2 kb bands; the F_1 was obtained by crossing 5 and 3 kb parent with 6 and 2 kb parent.

RFLP band (kb) combination	*Parental/recombinant type*	*Remarks*
6, 5, 3 and 2	Parental	Pattern of F_1 parent of back-cross progeny
6, 5 and 2	Recombinant	
6, 3 and 2	Recombinant	
6 and 2	Parental	Pattern of the parent to which back-cross was made

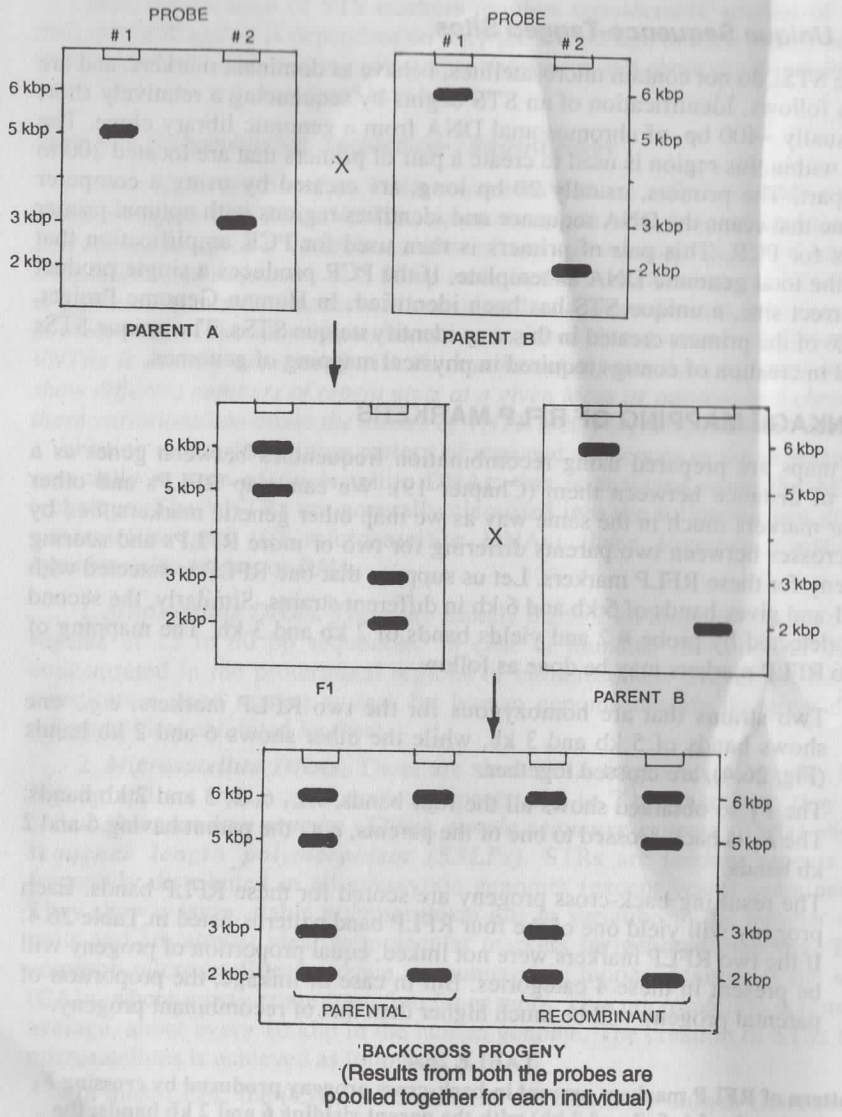


Fig. 26. 3. Results expected from a cross between two parents differing for two RFLP markers. In simple terms, the four types of backcross progeny are expected in the ratio of 1 : 1 : 1 : 1 if the two markers were not linked.

26.7.1. Estimation of Linkage

In actual analysis of RFLP data, linkage is detected by a statistical test called the ***lod*** (Logarithm of odds) ***score method*** developed by Morton in 1955. The theoretical basis of this method is quite complex, and computer programmes are used to analyze pooled data from a large number of pedigrees (human beings) or crosses. The programmes determine the probability that two markers exhibit a

certain degree of linkage, *i.e.*, are within a particular number of map units of distance, and also the probability that the given data would have been obtained if the two markers were unlinked. The lod score is then estimated according to the following formula.

$$\text{lod score} = \log_{10} \frac{\text{Probability of a certain degree of linkage}}{\text{Probability of independent assortment}}$$

If the lod score equals +3 or is greater, the two RFLP markers are considered to be linked. Since a $\log_{10}$ value of 3 equals 1,000, a lod value of 3 suggests that there is a 1,000-fold greater probability that the two markers are linked than they are assorting independently. If a lod score indicates the two RFLP markers to be linked, the number of recombinant offspring is divided by the total number of offspring to obtain the map distance in terms of map units or cM.

26.7.2. Preparation of Linkage Maps

RFLPs are quite common in virtually all species. Therefore, many RFLP markers can be analyzed for linkage relationships and a genetic map can be developed to depict the order and distances between pairs of a large number of different RFLP markers; such a map is known as an ***RFLP map.*** RFLP maps have become increasingly useful in the genetic analysis of plants and animals where genomes are large and relatively few genes have been mapped. RFLP maps can be used to locate functional genes within the genome, and they are also useful in physical mapping of eukaryotic genomes. Functional genes are located in RFLP maps by analysing linkage relationships of these genes with several RFLP markers, using the same approach that is applied for mapping of RFLP markers.

For a successful linkage mapping with DNA markers, sufficient DNA sequence polymorphism must be present between the parents used for hybridization. *Naturally outcrossing species tend to have high levels of DNA polymorphism, while the levels of DNA sequence variation are generally lower in naturally inbreeding species. Sometimes, mapping in inbreeding species requires that parents be as distantly related as possible.*

26.7.3. Mapping Populations

Several different kinds of populations can be used for linkage mapping of DNA markers. In the example described above (Fig. 26.4), a back-cross population was used. Other populations suitable for mapping are F_2, recombinant inbred lines (RILs) derived from a cross (easily developed by single seed descent procedure), and near-isogenic lines (NILs) developed by back-cross procedure. F_2 and back-cross populations are easy to construct, but they are ephemeral, *i.e.*, can not be maintained over a period of time, except by asexual propagation, and sterility in F_1 may limit some combinations of parents. Creation of RILs and NILs is time and labour-intensive (particularly in the case of NILs), and regions of the genome may tend to remain heterozygous longer than theoretically expected. Further, obligate outcrossing species are much more difficult to map using RILs because of the difficulty in selfing these species.

26.8. CREATION OF CONTIGS

First of all, chromosome sorting or some other strategy is used to obtain many copies of a single chromosome; the DNA from such a sample is used to construct a

chromosome-specific library. The next step in physical mapping is creation of a contig. A ***contig*** consists of a series of clones that contain overlapping pieces of DNA covering a specific region of a chromosome or even the entire chromosome (Fig. 26.4). Contigs are usually constructed using YAC and cosmid clones. The general approach in creation of contigs is to identify clones that have adjacent DNA segments from the chromosome; the members of contig must contain some overlapping regions to allow the precise determination of their location in the contig. The ultimate goal of physical mapping procedures is to obtain a complete contig for each chromosome of the genome.

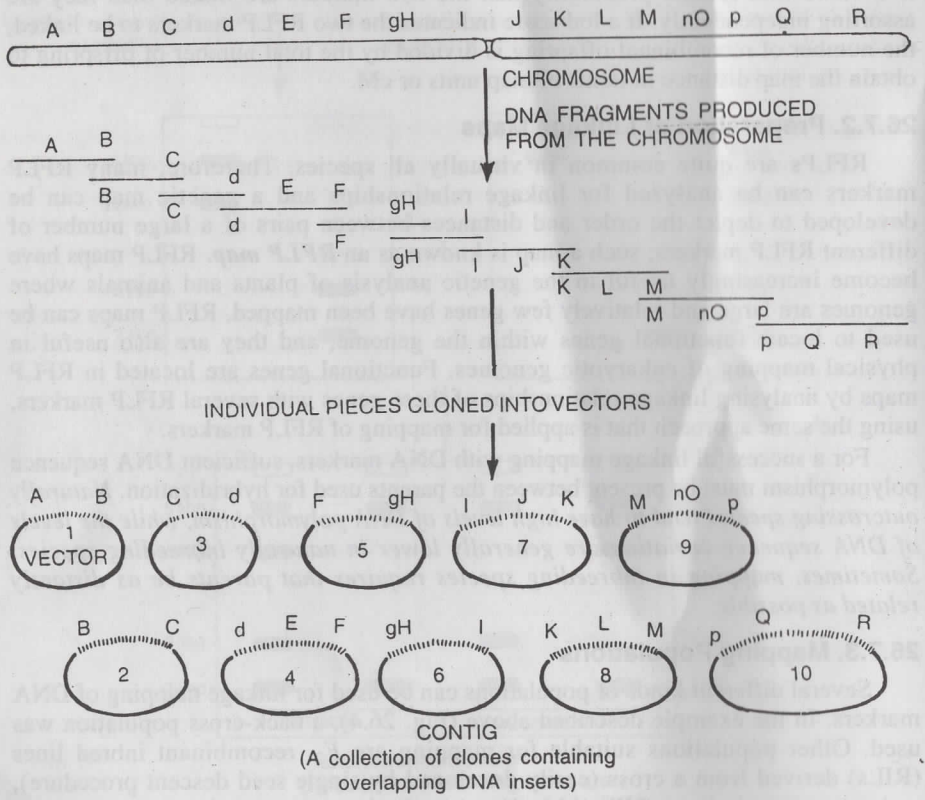


Fig. 26.4. A schematic representation of the concept of a contig. In practice, the progression of events in the creation of a contig is exactly the reverse of that shown here: overlapping series of clones are identified after cloning; this enables identification of the series of fragments that together show the entire chromosome.

A contig represents a physical map of a chromosome. The cloned DNA fragments of a contig can be correlated with locations along a chromosome obtained from linkage or cytogenetic mapping. This can be accomplished by identifying members of the contig that contain inserts having such genes that have already been mapped by linkage or cytological methods. This would permit the alignment of the other members of the contig along the chromosome. Unfortunately, for most eukaryotic species, few genes have been genetically mapped. Alternatively, RFLP

and other DNA markers (*e.g.*, STS) may be used to correlate the locations in a linkage map with the members of a contig.

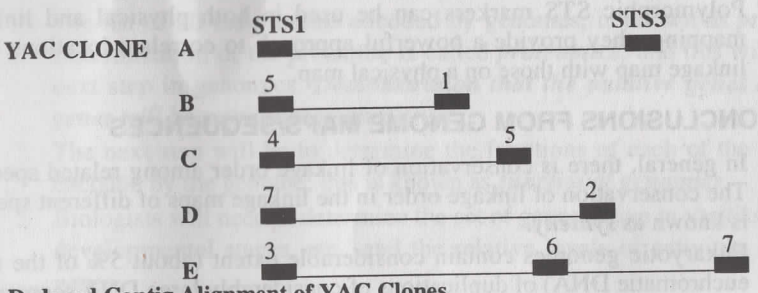
YAC cloning vectors are used to make contigs of eukaryotic chromosomes. YAC vectors can accept DNA inserts of several hundred kilobase pairs to perhaps 2 megabase pairs. Therefore, for small human chromosomes, a few hundred YACs would be sufficient to create a contig that would span the entire length of chromosome. In contrast, thousands or even tens of thousands of plasmid clones will be required for this purpose. Therefore, ***YAC cloning is usually the first step in creating a physical map of a genome***. But the large insert size of YAC clones makes them difficult to use in gene cloning and sequencing experiments. Therefore, ***DNA insert of each YAC clone of a contig is used to prepare a cosmid library***; a cosmid vector accepts a much smaller DNA insert (usually, tens of thousand base pairs in length) than do YAC vectors. In the Human Genome Project, bacterial artificial chromosome (BAC) libraries have been routinely created from YAC clones because BACs offer some unique advantages. Different experimental strategies can be employed to align the members of a contig. The strategy using unique STSs is briefly described below.

TABLE 26.5

Alignment of five YAC clones in a contig using seven unique STS markers

YAC clone	*STS Marker*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
A	+	–	+	–	–	–	–
B	+	–	–	–	+	–	–
C	–	–	–	+	+	–	–
D	–	+	–	–	–	–	+
E	–	–	+	–	–	+	+

STS Identity of YAC Clones



Deduced Contig Alignment of YAC Clones

YAC clone	*STS marker*						
	4	*5*	*1*	*3*	*6*	*7*	*2*
C	4	5					
B		5	1				
A			1	3			
E				3		7	
D						7	2

26.8.1. Creation of Contigs Using Unique STSs

Unique STSs can be used to identify and align clones of a contig. We take a simple example based on seven different unique STS markers identified by sequencing short segments of DNA from a chromosome 6-specific library. These 7 STSs are found in chromosome 6, but their order is unknown. Five different YAC clones from a chromosome 6-specific library were screened for the presence of these 7 STS markers, and the data listed in Table 26.5 were obtained. The presence of a STS marker in a YAC clone will be shown by positive PCR amplification using the primers specific for that STS marker; there will be no amplification if the STS were absent in the clone. In our example, most of the YAC clones were positive for two STS markers. The presence/absence of STS markers can be used to align the five YAC clones in the proper order to generate a contig for a region of chromosome 6 (it is expected to have many more YAC clones in a contig spanning the whole chromosome).

26.9. CORRELATING CYTOGENETIC, LINKAGE AND PHYSICAL MAPS

Cytogenetic, linkage and physical maps can be correlated using one of the following four approaches.

1. *In situ* hybridization is a common method in which YAC or cosmid clones can be hybridized to intact chromosomes to determine their location. This approach allows correlation of a cytogenetic map with a physical map.
2. In case of those genes that have already been cloned, linkage map can be correlated with a cytogenetic map by using the cloned gene as a probe in *in situ* hybridization.
3. A linkage map can be correlated with a cytogenetic map when a genetically mapped gene can be localized to particular chromosomal region in man-mouse somatic cell hybrids.
4. Polymorphic STS markers can be used in both physical and linkage mapping; they provide a powerful approach to correlate locations on a linkage map with those on a physical map.

26.10. CONCLUSIONS FROM GENOME MAPS/SEQUENCES

1. In general, there is conservation of linkage order among related species. The conservation of linkage order in the linkage maps of different species is known as ***synteny.***
2. Eukaryotic genomes contain considerable extent (about 5% of the total euchromatic DNA) of duplications of considerably large DNA segments.
3. There is evidence for movement of DNA from mitochondria/chloroplasts into the nuclear genomes.
4. Human genome is estimated to contain about 31,000 to 39,000 genes.
5. In human genome, over 50% of the DNA consists of repeated sequences: 45% of the DNA is made up of four classes of parasitic (transposon) DNA elements (virtually all of them are enfeebled).
6. The genome has G.C–rich and A.T–rich regions; the genes are more densely distributed in G.C–rich regions and their average intron size is smaller than those in the A.T–rich regions.

7. Of the 1,278 protein families encoded by only ~5% of the human genome, only 94 families are specific to vertebrates.
8. Hundreds of human genes seem to have come directly by horizontal transfer from bacteria, rather than having evolved from bacterial genes.
9. The pericentromeric and subtelomeric regions of chromosomes are filled with large recent segmental duplications of sequences from elsewhere in the genome.
10. The mutation rate is twice as high in male as in female meiosis; thus ***most mutations occur in males***.
11. Cytogenetic analysis of the sequenced clones confirms that large G.C–poor regions are strongly correlated with 'dark G–bands' in karyotypes.
12. Recombination rates tend to be much higher in distal regions (around 20 megabases) of chromosomes and on shorter arms in general.
13. More than 1-4 million single nucleotide polymorphisms (SNPs) have been identified in the human genome.

26.11. BEYOND THE GENOME SEQUENCE

By February 2001, when conclusions from the human draft sequence were published, complete sequences of genomes of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, 7 archaea, one fungus, two animals and one plant had become available. Once the complete base sequence of a genome becomes available, scientists can begin to unravel the total information contained in the genome; this is termed as ***functional genomics.*** Models then can be developed that would explain how biological systems develop and respond using this information. The following activities, therefore, may be expected to predominate the genomics science after the genome sequence becomes available.

1. The list of all the proteins encoded by genomes, is known as ***proteome.*** Determination of the proteome is called ***proteomics,*** and this will be the next step in genomics. ***Demonstration that the putative genes are real genes will be a continuing challenge.***
2. The next step will be to determine the functions of each of the proteins encoded by the genome; this is known as ***functional genomics.***
3. Biologists will need to determine the set of genes active in various tissues, developmental stages, etc., and the relative levels of activities of these genes.
4. Determination of the complete set of functional sequences, other than genes, *e.g.,* regulatory sequences, present in a genome will be much more difficult than that of genes themselves.
5. The most challenging task will be to develop models, based on the proteins encoded by specific genomes, that account for the development and responses of biological systems and phenomenon, and to evaluate the validity of these models.
6. The field of genetic epidemiology is expected to make rapid progress after genome sequencing. ***Genetic epidemiology*** is the analysis in large populations of the interactions of gene, genetic variation and environment;

this allows the discovery of new drug treatments, and to tailor them to individual genetic and life profiles. Any two unrelated human beings differ from each other by one base pair in every ~1,000 bp of their genomes; these differences are called ***single-nucleotide polymorphisms*** or ***SNPs*** (often called '*snips*') since the differences are of one base pair only. About 85% of all difference in human DNAs are SNPs; it is generally considered that SNPs are probably the cause of most common genetic disorders. It is proposed to determine the 3,00,000 of the commonest SNPs in human populations, and to use this information to design new drug treatments for the common disorders.

26.12. APPLICATIONS OF MOLECULAR MARKERS

DNA markers can be used for the following different purposes.

1. They are used for preparation of linkage maps just as are NEP/protein markers (Section 26.7).
2. Special mapping strategies can be used for rapidly targetting specific regions of the genome.
3. DNA markers can be used for map-based cloning of genes of interest.
4. RFLP markers mapped in one species or genus can often be used to construct parallel genetic maps in related species or genera; this is called ***parallel mapping***.
5. DNA markers have been used to map quantitative trait loci (QTLs).
6. DNA markers can be used for indirect selection for difficult to select traits like yield, etc.; this is known as ***marker-assisted selection*** (***MAS***).
7. DNA markers can be used for rapid and unequivocal diagnosis of human diseases.
8. DNA markers are used for assessment of genetic diversity, taxonomic classification and for DNA fingerprinting used for identification of individuals/varieties/criminals/parents, etc.

QUESTIONS

1. What are genome maps and genetic maps ? Briefly describe the methods for preparation of cytogenetic maps, and the approaches for correlating cytogenetic, linkage and physical maps.
2. List the various types of genetic markers, describe in some detail their various features and compare their attributes.
3. What are DNA markers ? Give a brief description of some important DNA markers and discuss their applications.
4. Briefly describe the approach for developing sequence-tagged site markers, and explain their uses in linkage and physical mapping.
5. Briefly describe the aims, strategies and achievements of the human genome project.
6. Briefly describe the linkage mapping of DNA markers, and the strategy for map-based cloning.
7. Briefly describe the chief strategies used for developing chromosome-specific libraries.
8. What is a contig ? Describe how the contigs are constructed and used in genome mapping ?
9. Briefly describe the strategy of physical mapping of a genome.

10. 'Much remains to be done even after a genome has become completely sequenced.' Comment.
11. Write short notes on:

(*i*) Genomics, (*ii*) functional genomics,
(*iii*) genetic marker, (*iv*) molecular marker,
(*v*) DNA marker, (*vi*) STS,
(*vii*) NEP marker, (*viii*) linkage map,
(*ix*) genetic map, (*x*) cytogenetic map,
(*xi*) physical map, (*xii*) Human Genome Project,
(*xiii*) RFLP, (*xiv*) RAPD,
(*xv*) sequence tagged sites, (*xvi*) polymorphic STS,
(*xvii*) unique STS.

27

Molecular Techniques in Genetics

27.1. INTRODUCTION

The questions and quests in genetic studies have become increasingly focussed on molecular events and the mechanisms involved therein. As a consequence, a host of molecular techniques have been developed; these techniques are highly sensitive, unbelievingly accurate and often surprisingly rapid. In general, the information generated from these techniques provides the bulk of our knowledge about genes, their function, regulation, etc. Therefore, some appreciation of the most common of such techniques becomes essential in order to be able to fully appreciate the discussions on various aspects of genes. These techniques have been divided into the following eight arbitrary groups:

(1) *nucleic acid reassociation kinetics,*
(2) *physical mapping of chromosomes* (heteroduplex mapping, restriction mapping),
(3) *nucleic acid hybridization* (*in situ* hybridization, Southern hybridization, Northern hybridization, dot blot technique, colony hybridization, chromosome walking),
(4) *labelling of nucleic acids* (nick translation),
(5) *DNA foot-printing,*
(6) *molecular markers* (DNA fingerprinting, RFLP maps, RAPDs maps),
(7) *DNA synthesis* and
(8) *DNA sequencing.*

The treatment of these techniques, of necessity, is in general terms as the aim is only to acquaint the reader with them.

27.2. NUCLEIC ACID REASSOCIATION KINETICS

Studies of DNA reassociation kinetics have provided information on the existence of repetitive DNA sequences as well as on the number of copies per genome of such sequences. These studies have permitted the classification of repetitive sequences into highly repetitive (>10^5 copies per genome) and moderately repetitive (2 to <10^5 copies/genome) DNA (see, Chapter 25). In such a study, a sample of DNA is generally sheared (broken) into fragments of small size (say of 400-500 bp). It is then subjected to high temperature or high pH, which leads to the separation of complementary strands producing single-stranded DNA; this is known as ***denaturation.*** When a denatured DNA sample is allowed to cool (in case of heat denaturation) or its pH is lowered (in case of alkali denaturation), the complementary single strands reassociate through complementary base pairing to produce double-stranded DNA; this phenomenon is called ***renaturation.***

If RNA is added into a sample of denatured DNA, RNA strands will associate with their complementary DNA strands producing DNA : RNA hybrid molecules; this is referred to as ***DNA·RNA hybridization.*** If, however, two samples of denatured DNA are mixed, DNA : DNA hybrid molecules will be formed; such a process is known as ***DNA hybridization.*** Both these phenomena taken together constitute the technique of ***nucleic acid hybridization.***

Renaturation rate in a denatured sample of DNA is mainly a function of DNA concentration and time allowed for renaturation; these two parameters are expressed as *Cot* (initial concentration × time), which is estimated as moles per litre × seconds as follows.

$$\frac{C}{C_0} = \frac{1}{1 + k.Cot}$$

where, C_0 and C are the concentrations in moles of nucleotides per litre at the initial time t_0 and after reassociation reaction has progressed for time t, respectively, and k is a reassociation rate constant. Thus the parameter controlling reassociation reaction is the product of initial concentration of single-stranded DNA (C_o) and the time of incubation for reassociation (t); this is simply described as the *Cot*. The rate of reassociation reaction is governed by the equation given below.

$$\frac{dC}{dt} = -kC^2$$

The rate of renaturation of a given DNA segment is proportional to the number of copies of that segment present in a unit DNA solution; the more the number of copies, the faster would be its renaturation. This is because renaturation would occur when a DNA strand collides or comes very close to its complementary strand. Since such a collision depends entirely on chance, the chances of collision will be proportional to the number of copies of the concerned segment present in a given solution.

In renaturation studies, samples are withdrawn at regular intervals from a DNA sample undergoing renaturation and the proportion of double-stranded DNA in the sample may be measured in one of the following three ways. (1) Single-stranded DNA has a higher optical density (OD) at a specific wavelength than does double-stranded DNA. Therefore, OD of a renaturating DNA sample would decline with time, and the change in OD may be used to estimate the amount of double-stranded DNA in the sample. (2) The DNA sample may be passed through a column of hydroxyapatite, which retains double-stranded DNA but allows single-stranded DNA to pass through. The amount of DNA retained by the column may then be easily determined. (3) The sample may be treated with a specific nuclease say, S1 nuclease, which digests only single-stranded DNA. The DNA left in the sample after such a digestion will be double-stranded, and its quantity can then be estimated.

A ***Cot curve*** is prepared by plotting the fraction of DNA that has reassociated [= $1 - (C/C_0)$] against the log of *Cot* values. The *Cot* curves for poly-U : poly-A (complexity = 1 bp), genomes of phages MS2 (complexity = 3,500 bp) and T4 (complexity = 1.7×10^5 bp), and of *E. coli* (complexity = 4.2×10^6) follow similar pattern, *viz.*, sigmoidal, and the ratio of *Cot* values between 10% and 90% reactions is ~100 in each case. However, the *Cot* required in each case is different.

The renaturation curve of eukaryotic DNA is quite complex as compared to that of *E. coli* DNA. The curves for eukaryotic DNAs show three distinct regions

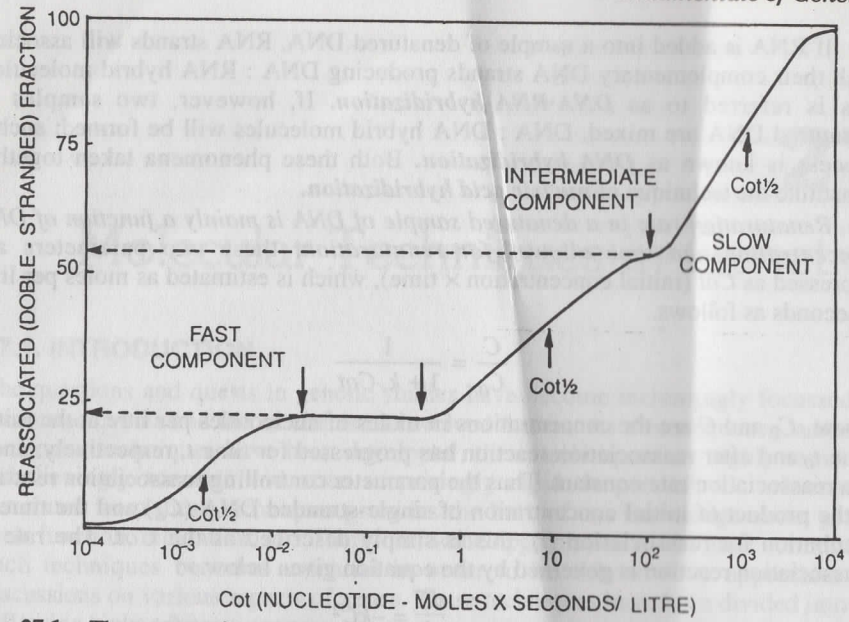


Fig. 27.1. The renaturation pattern of an eukaryotic DNA sheared into fragments of about 400 nucleotides. The earliest renaturation occurs in the highly repetitive DNA, followed by moderately repetitive DNA and, at last, unique sequences.

representing highly repetitive, moderately repetitive and nonrepetitive or unique DNA sequences (Fig. 27.1). Highly repetitive sequences are the earliest to renature (***fast component*** in Fig. 27.1) since they have very large number of copies per genome. Moderately repetitive sequences are the next to reassociate (***intermediate component***), while the unique sequences renature towards the end (***slow component***).

Cot curve provides information on the proportions of fast, intermediate and slow components, which is about 24%, 31% and 45%, respectively, in Fig. 27.1. In addition, the *Cot* values for the 3 components, which are approximately 2×10^{-3}, 8 and 10^{3}, respectively, can be used to estimate the complexity and the copy number for each of these components.

27.3. PHYSICAL MAPPING OF CHROMOSOMES

Physical maps of chromosomes depict such recognition features as the sites for restriction enzymes (see later), sites and sizes of deletions, sites of A·T-rich regions, etc. of the genetic material, and locations of these features are shown in terms of the physical distance (in bp) between them. Physical maps may also depict genes in the same manner. In contrast, when recombination data is used for constructing the maps of chromosomes, it is known as ***linkage map.*** Here two main types of physical mapping, *viz*, heteroduplex mapping and restriction mapping, are briefly described.

27.3.1. Heteroduplex Mapping

A ***heteroduplex*** is a double-stranded nucleic acid molecule the two strands of which are not entirely complementary. For example, in a DNA heteroduplex one strand may contain one allele of a gene, while the other strand may have another

allele. Such heteroduplex molecules are prepared by mixing two denatured DNA preparations (*e.g.*, DNA from a deletion mutant of the r^{II} locus of T_4 phage and that from the wild type phage) and allowing them to renature. Many of the double-stranded molecules thus produced will be heteroduplexes, *i.e.*, they will be composed of one strand from the wild type phage DNA and the other strand from the deletion mutant of the r^{II} locus. In such heteroduplexes, the segment of wild type strand, which is missing in the mutant strand will remain unpaired and can be seen as a loop under an electron microscope (Fig. 27.2).

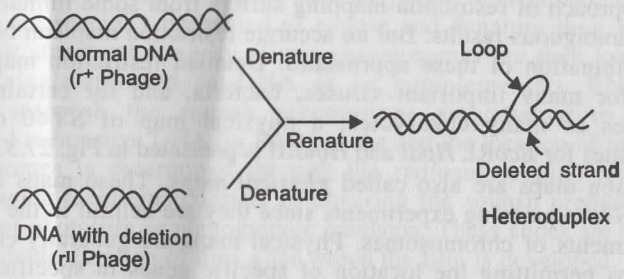


Fig. 27.2. A heteroduplex DNA molecule showing the deleted segment extending as a loop from the normal DNA strand. This heteroduplex was produced by pairing between a normal (r^{+}) and its complementary strand having deletion (r^{II}) in the case of r^{II} deletion mutants of phage T4.

This technique is known as ***heteroduplex mapping.*** It permits determination of the size and location of a deletion in a gene. Using this technique, the presumed deletion mutants of r^{II} locus by Benzer were shown to really contain deletions in the r^{II} region. Further, the size of r^{II} locus was estimated as 2645 bp (cistron A, 1800 ± 70 bp; and cistron *B*, 845 ± 50 bp). This technique is extended to DNA : RNA heteroduplexes, which has permitted the discovery and demarcation of introns and exons of eukaryotic genes (see, Chapter 25).

27.3.2. Restriction Mapping

A ***restriction map*** of a chromosome (or a segment thereof) depicts the sites where one or more restriction endonucleases produce cuts in the chromosome. Several approaches are used for restriction mapping, *e.g.*, comparison of the products of (*i*) partial digestion with one enzyme; (*ii*) sequential digestion with two or more different restriction enzymes; (*iii*) digestion of a circular DNA molecule under conditions permit that only one cut by an endonuclease (in place of its complete digestion), followed by digestion of the linear segments thus produced by another endonuclease, which induces only one cut at a specific site in the chromosome; and several other approaches.

In the ***partial digestion approach,*** one end of the linear DNA molecule is labelled using ^{32}P-ATP and the enzyme polynucleotide kinase or through some other approach. These molecules are now partially digested with the desired endonuclease so that it induces on an average one cut per DNA molecule. The partial digest is subjected to gel electrophoresis and the positions of the labelled bands are determined by autoradiography. The fastest band thus detected will be the fragment produced by cleavage at the recognition site nearest to the labelled end of the DNA

molecule. The next fastest band will have the fragments produced by cleavage at the second recognition site from the labelled end, and so on. The slowest band will contain the original DNA molecule. The size of DNA fragments present in the different labelled bands can be determined as follows. A control lane is run in the same gel; this lane has standard DNA fragments of different but known sizes. The locations of bands in the lane for partial digets are compared with those in the control lane; this provides information on the sizes of various fragments present in the partial digest. This information is then used to prepare a restriction map of the DNA molecule.

Each approach of restriction mapping suffers from some limitations and may yield some ambiguous results. But an accurate restriction map can be prepared by using a combination of these approaches. Detailed restriction maps have been developed for many important viruses, bacteria, and for certain sections of chromosomes of many eukaryotes; a physical map of SV40 depicting the recognition sites for *Eco*RI, *Hpa*I and *Hin*dIII is presented in Fig. 27.3.

Restriction maps are also called ***physical maps.*** These maps are extremely useful in DNA sequencing experiments since they are helpful in the production of desired fragments of chromosomes. Physical maps are generally combined with genetic maps permitting the location of specific genes in specific segments of chromosomes.

Electron micrographs of partially denaturated chromosomes may be compared with the denaturation data of different DNA fragments generated through endonuclease digestion. The single-stranded regions in such electron micrographs would represent A : T-rich regions of chromosomes; fragments corresponding to such regions would denature more readily than those corresponding to the double-stranded regions of chromosomes. In this manner, specific DNA fragments may be assigned to specific areas of chromosomes.

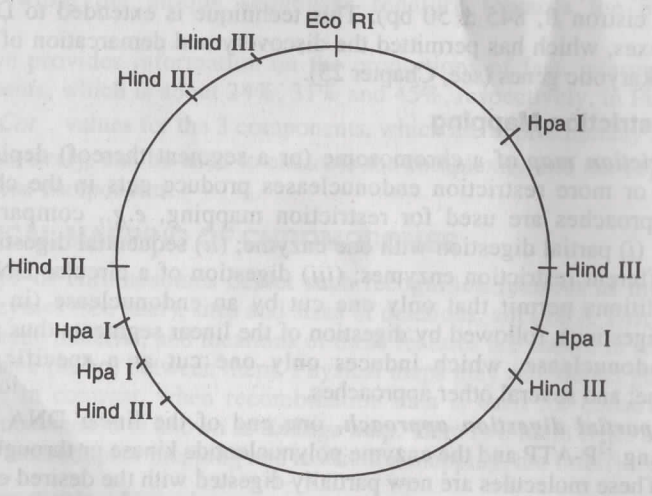


Fig. 27.3. A physical map (restriction map) of the Simian Virus 40 (SV40) chromosome showing the recognition sites for three restriction endonucleases, *viz.*, *Eco*RI, *Hpa*I and *Hin*dIII. The recognition site for *Eco*RI has been arbitrarily placed; other recognition sites are mapped in relation to this recognition site.

27.4. NUCLEIC ACID HYBRIDIZATION

The phenomenon of complementary base pairing between two DNA strands, and between one DNA and one RNA strand having similar base sequences has been exploited to develop a very large number of sensitive techniques; some of the common techniques are briefly described here.

27.4.1. *In Situ* Hybridization

In situ *hybridization* has been used to establish the location of sat-DNA in chromosomes. For this purpose, radioactive copies of sat-DNA or its complementary RNA (using sat-DNA as templete) are prepared. Chromosomes in squash preparations are specially pretreated to expose and denature their DNA without affecting their structural integrity. These chromosomes are then loaded with the radioactive single-stranded sat-DNA or its complementary RNA. After an appropriate interval, the squash preparations are washed to remove the nonhybridized radioactive DNA (or RNA) probe and the location of radioactivity in the chromosomes is determined through the radioautographic technique. The ***radioautographic technique*** consists of covering the squash preparations with a special type of photofilm or photographic emulsion. These slides are then stored in dark for a suitable duration of time after which the film is developed. Care is taken to avoid any disturbance of the photographic film placed on the squash preparations. When such a squash preparation is now viewed under light microscope, the areas having radioactivity would show silver grains on the photographic film (due to the radioactive decay). Using this technique, it was shown that sat-DNA is mainly located in the centromeric and telomeric regions of chromosomes.

In situ or ***cytological hybridization*** is also used to locate specific genes in chromosomes, especially in giant chromosomes. For this purpose, a radioactive clone representing a gene, most often labelled cDNA copies of the mRNA produced by the gene, is used as probe. This approach has enabled the location of specific sequences in particular bands of salivary gland chromosomes of *Drosophila*. Colony hybridization (see, later) is also a form of *in situ* hybridization.

This technique is also useful in disease diagnosis, particularly for the detection of viruses in tissues and cells. In simple terms, the tissues are fixed in formalin, embedded in paraffin wax, sectioned, stained with a suitable stain, *e.g.*, haematoxylin and eosin, and the slides are viewed under light microscope. The sections are then exposed to specific nonradioactive probes and viewed again under the light microscope. This permits the correlation of conventional microscope data with that obtained with the probes and to determine the location of the concerned pathogens in the tissues and cells.

27.4.2. Southern Hybridization

The name of this technique is derived from the following: (1) the name of its inventor, E.M. Southern and (2) the DNA : DNA hybridization that forms its basis. It is also called ***Southern blotting*** since the procedure for transfer of DNA from the gel to the nitrocellulose filter resembles blotting. This technique has since been extended to the analysis of RNA (***northern blotting***) and proteins (***western blotting***); but these names are only jargon terms, *i.e.*, reverse of Southern being northern and so on, and do not reflect any functional or historical significance.

In Southern hybridization, a sample of DNA containing fragments of different sizes is subjected to electrophoresis using either polyacrylamide or agarose gel. The

DNA sample may either be subjected to mechanical shearing or to restriction endonuclease digestion in order to generate the fragments. Agarose gel is useful in separating DNA fragments of few hundred to 20 kb in size, while polyacrylamide is preferred for smaller fragments. Very large DNA fragments of upto 1000-2000 kb are separated in agarose gel with pulsed electrical fields, or field inversion.

The gel provides a complex network of polymeric molecules through which DNA fragments migrate, depending on their sizes, under an electric field since DNA molecules are negatively charged. Smaller molecules of DNA migrate relatively faster than do the larger ones. Marker DNA fragments of known sizes are run in a separate lane; this permits an accurate determination of size of an unknown DNA molecule by interpolation. The gels are stained with intercalating dye ethidium bromide, which gives visible fluorescence on illumination of the gel with UV light; as little as 0.05 μg of DNA in one band can be detected by using this dye. This approach is useful when few DNA fragments with considerable differences in their lengths are to be separated and studied. This approach also separates the closed circular (supercoiled), nicked (relaxed) and linear configurations of a single DNA molecule.

In many situations, it is critical to detect and identify DNA fragments in a sample that are complementary to a given DNA sequence, *e.g.*, to demonstrate the presence of the gene in question in transgenics, to detect and study RFLP (restriction fragment length polymorphism; see, later), etc. This is achieved by Southern hybridization in which the following steps are performed.

1. The restriction fragments of DNA present in the agarose gel (after electrophoresis) are denatured into single-stranded form by alkali treatment.
2. They are then transferred onto a nitrocellulose filter membrane; this is done by placing the gel on top of a buffer saturated filter paper, then laying the nitrocellulose filter membrane on top of the gel, and finally placing some dry filter papers on top of this membrane. The buffer moves, due to capillary action, from the bottom filter paper to that on the top through the gel carrying with it the denatured DNA present in the gel; the DNA becomes trapped in the nitrocellulose membrane as the buffer pases through it. This process is known as ***blotting*** and takes several hours to complete. The relative positions of the bands on the membrane remain the same as those in the gel and there is a minimal loss in their resolution (sharpness).
3. The nitrocellulose membrane is now removed from the blotting stack, and the DNA is permanently immobilized on the membrane by baking it at 80°C *in vacuo*.
4. Single-stranded DNA has a high affinity for the nitrocellulose filter membrane. (Note that RNA lacks this affinity). Therefore, the baked membrane is treated with a solution containing 0.2% each of Ficoll (an artificial polymer of sucrose), polyvinylpyrrolidone and bovine serum albumin; this mixture is often supplemented with an irrelevant nucleic acid, *e.g.*, t-RNA (***pretreatment***). This treatment prevents nonspecific binding of the radioactive probe (to be used in the next step) probably by attaching macromolecules to all the free binding sites on the membrane. Often the above mixture is included in the hybridization reaction itself.

5. The pretreated membrane is placed in a solution of radioactive, single-stranded DNA or an oligodeoxynucleotide (a DNA segment having few to several nucleotides) called ***probe***. The name probe signifies the fact that this DNA molecule is used to detect and identify that DNA fragment in the gel/membrane, which is complementary to the probe. The conditions during this step are chosen so that the probe hybridizes with the complementary DNA on the membrane to the greatest extent with a low nonspecific binding on the membrane; this step is known as ***hybridization reaction***. Usually the initial hybridization reaction is carried out under conditions of relatively low stringency of hybridization to permit a high rate of hybridization.
6. After hybridization reaction, the membrane is washed to remove the unbound probes. Usually, the hybridization is followed by a series of post-hybridization washes of increasing stringency, *i.e.*, higher temperature or, more commonly, lower ionic strength, with a view to eliminate pairing of the radioactive probe to related sequences and to allow only perfectly complementary pairing.
7. The membrane is now placed in close contact with an X-ray film and incubated for a desired period to allow images due to the radioactive probes to be formed on the film. The film is then developed to reveal distinct band(s) corresponding to the positions in the gel of the DNA fragments that are complementary to the radioactive probe used in the study.

It should be kept in mind that electrophoresis of sheared or restricted DNA produces a smear in which the fragments are distributed in a continuoum according to their size and that there are no distinct bands as such. The distinct bands on the photographic film are produced by the hybridization reaction of the selected probe with one or few fragments present in the gel.

The Southern blotting technique is extremely sensitive. It can be used to map the restriction sites around a single copy gene sequence in any genome (even of man). It is used for DNA fingerprinting (see, later), preparation of RFLP maps, detection and identification of the transferred genes in transgenic individuals, etc.

Recently, some new membrane materials, *e.g.*, nylon membranes, have been developed, which have the following advantageous features: (1) they are physically robust in comparison to nitrocellulose filter membranes, (2) both DNA and RNA become cross-linked to them by a brief exposure to UV light, which (3) saves the time needed for baking *in vacuo* in the case of nitrocellulose membranes, and (4) the same membrane blot, *e.g.*, a membrane onto which DNA/RNA has been transferred from a gel and cross-linked by UV exposure, can be used for search with more than one probe after removing the earlier probe by high temperature washing or some other denaturing procedure; in other words, the nylon membranes are reusable.

27.4.3. Northern Hybridization

In this technique, RNAs are separated by gel electroporesis, the RNA bands are transferred onto a suitable membrane, *e.g.*, diazobenzyloxymethyl (DBM) paper or nylon membranes, and immobilized; the bands are hybridized with radioactive single-stranded DNA probes, and the bands showing hybridization are detected by autoradiography.

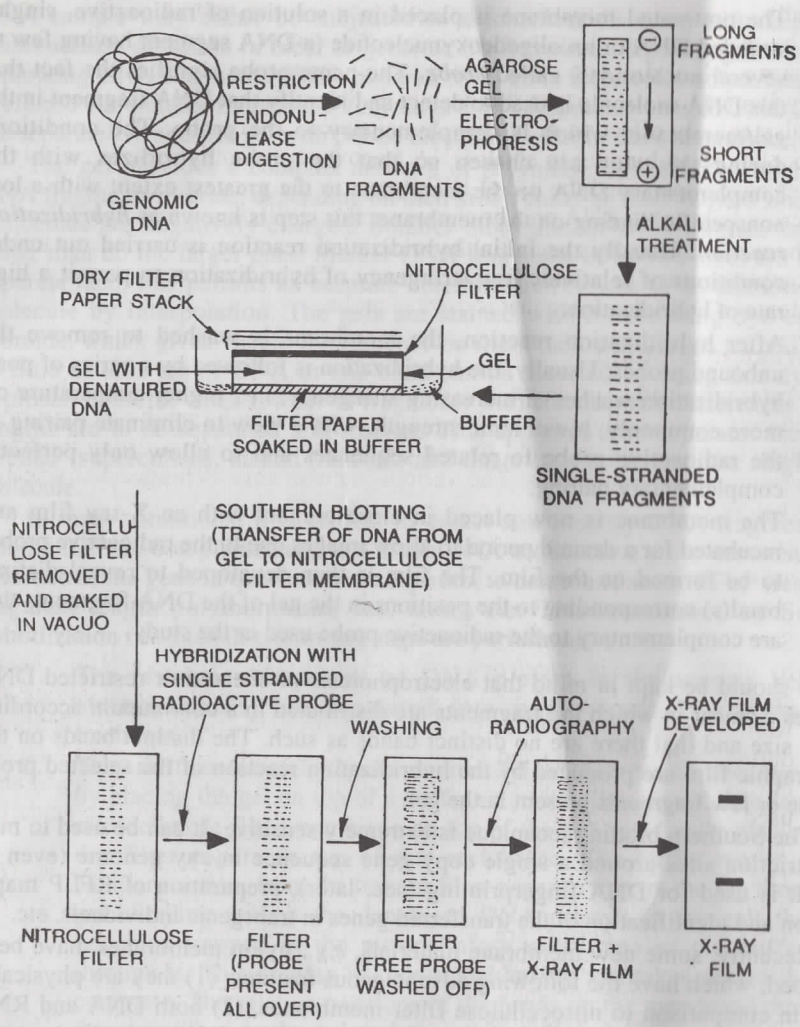


Fig. 27.4. A schematic representation of the steps involved in separation of DNA fragments in a gel, their blotting onto a nitrocellulose membrane filter and detection of a specific sequence through hybridization with a radioactive single-stranded probe.

Clearly, northern blotting hybridization is simply an extension of the Southern blotting technique (Fig. 27.4); the two techniques, however, show the following differences: (1) in Southern hybridization *DNA is separated* by gel electrophoresis, while in northern blotting *RNAs are separated*; (2) as a result, in Southern hybridization DNA has to be denatured before blotting, while this step is not needed in northern hyridization; (3) nitrocellulose membrane is generally not used for northern, while it is often used for Southern hybridization; and finally, (4) hybridization with the probe produces DNA : DNA hybrid molecules in Southern but RNA : DNA molecules in northern hybridization.

Initially, specially prepared paper (diazobenzyloxymethyl, DBM, paper prepared by diazotization of aminobenzyloxymethyl paper) was used for northern blotting since RNA did not bind to nitrocellulose membrane. RNA becomes covalently bound to DBM paper due to which these blot-transfers are reusable. DBM is also equally effective in binding to denatured DNA, and is more efficient than nitrocellulose in binding to small DNA fragments. Recently developed nylon membranes have superseded the use of DBM paper as they are robust, reusable and bind (by cross-linking) to RNA on a brief exposure to UV light.

Northern hybridization is useful in the identification and separation of the RNA that is complementary to a specific DNA probe; this is a sensitive test for the transcription of a DNA sequence that is used as probe.

27.4.4. Western Blotting

In western blotting, proteins are electrophoresed in polyacrylamide gel, transferred onto a nitrocellulose or nylon membrane (to which they bind strongly), and the protein bands are detected by their specific interaction with antibodies, lectins or some other compounds. The various steps of this technique are briefly described below.

1. Proteins are separated by polyacrylamide gel electrophoresis.
2. The protein bands are transferred onto a nitrocellulose or nylon membrane; initially this was achieved by a capillary movement of buffer similar to Southern blotting (***capillary blotting***). But now-a-days, it is usually done by electrophoresis (***electrophoretic blotting***). Electrophoresis has been applied for the blotting step in Southern and northern hybridizations as well; in such cases, buffer of low ionic strength (to avoid overheating during electrophoresis) and nylon membranes (since nucleic acids bind to nitrocellulose membrane only under conditions of high ionic strength) are used. The electrophoretic blotting, both of proteins and nucleic acids is much faster and more efficient than capillary blotting.
3. The specific protein bands are identified in a variety of ways. (*i*) Antibodies are the most commonly used as probes for detecting specific antigens. (*ii*) Lectins are used as probes for the identification of glycoproteins. These probes may themselves be radioactive or a radioactive molecule may be tagged to them. Often the identification process is based on a ***'sandwich' reaction.*** In such an approach, a species-specific second antibody or protein A of *Staphylococcus aureus* (protein A binds to certain subclasses of IgG antibodies) or streptavidin (it binds to biotinylated antibodies) is used to bind to the antibodies bound to the protein bands. These second molecules may be labelled with radioactive, enzyme or fluorescent tags; a single preparation of these labelled molecules can be employed as a general detector for various probes.

27.4.5. Dot Blot Technique

This technique is used to detect the presence of a given sequence of DNA/RNA in the nonfractionated (not subjected to gel electrophoresis) DNA; sample DNA's from several tissues/individuals can be tested in a single test-run. Dot blots are useful in detecting the presence of the sequence being transferred in a number of suspected transgenic individuals, the presence of the specific mRNA in several such

individuals or in different tissues of a single individual. The generalized procedure of dot blot is briefly described below.

1. The sample DNAs (or RNAs) from different individuals or tissues are transferred onto a nitrocellulose filter in form of dots; several samples are dot-blotted onto a single filter (Fig. 27.5). A sample DNA or RNA is the total DNA or RNA extracted from one individual or tissue.
2. The DNA is first denatured and then the filter is baked at 80°C to fix the DNA firmly onto the filter. The filter is then pretreated appropriately to prevent nonspecific binding of the probe to the filter (see, Southern hybridization).
3. The filter is then treated with the appropriate radioactive single-stranded DNA probe under conditions favouring hybridization. The filter is then washed repeatedly to remove the free probe.
4. The dots having the appropriate DNA or RNA sequence will hybridize with the radioactive probe. These dots are detected by autoradiography; the intensity of a dot in the autoradiograph corresponds fairly well with the extent to which DNA or RNA is represented in the concerned clone.

The dots that show up in the autoradiograph denote the individuals or tissues in which the DNA or RNA sequence corresponding to the probe is represented. In case of transgenic individuals, Southern hybridization may then be used for the dot blot positive individuals to derive more precise information on whether and where the sequence in question is integrated into the genome.

27.4.6. Colony Hybridization

This technique is used to identify those bacterial colonies in a plate, which contain a specific DNA sequence. These bacterial colonies are obtained from bacterial cells into which this sequence was introduced through genetic transformation, and the given sequence is represented by the probe used in the hybridization experiment. The procedure for colony hybridization is briefly described below.

1. The bacterial cells subjected to transformation (*i.e.*, introduction of the vector containing a given DNA insert into bacterial cells) are plated on to a suitable agar plate; this is the ***master plate.***
2. The colonies of the master plate are replica-plated on to a nitrocellulose filter membrane placed on agar medium. For ***replica plating***, a block of wood or cork, suitable for the master plate is covered with velvet cloth. This block is sterilized and then lowered into the master plate till the velvet touches all the colonies; the block is withdrawn and gently lowered onto the nitrocellulose filter so that the bacterial cells sticking on to the velvet are transferred onto the filter. The master plate is retained intact for later use. A reference point is marked both on the master plate and on the replica plate to facilitate later comparisons.
3. After colonies appear on the filter, the filter is removed from agar plate and treated with alkali to lyse the bacterial cells. This also denatures the DNA of these cells.
4. The filter is treated with proteinase K to digest and remove the proteins; the denaturated DNA remains bound to the filter.

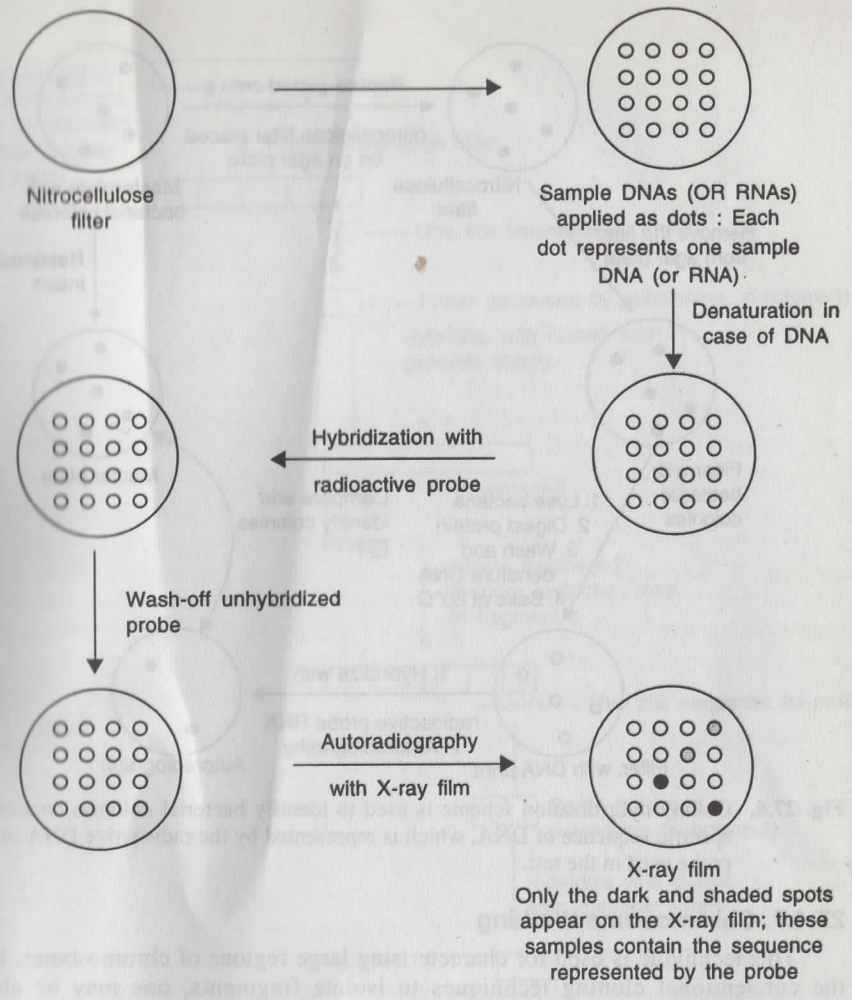


Fig. 27.5. A schematic representation of dot blot technique for detecting the presence of a specific sequence in the DNA (or RNA) of several individuals or tissues.

5. The filter is now baked at 80°C to fix the DNA onto the filter; this yields DNA-print of the bacterial colonies in the same relative positions as those of the colonies themselves in the master plate.
6. The filter is now hybridized with radioactive probe DNA or RNA; the probe represents the sequence of DNA segment used for transformation. The unhybridized probe is removed by repeated washing.
7. The colonies whose DNA hybridizes with the probe are detected by autoradiography; only these colonies show up in the autoradiograph.

The position of colonies showing up in the autoradiograph is compared with the masterplate to identify these colonies; these colonies contain the DNA segment in question. These colonies are then picked up for further studies. A variation of this procedure can be applied to phage plaques as well.

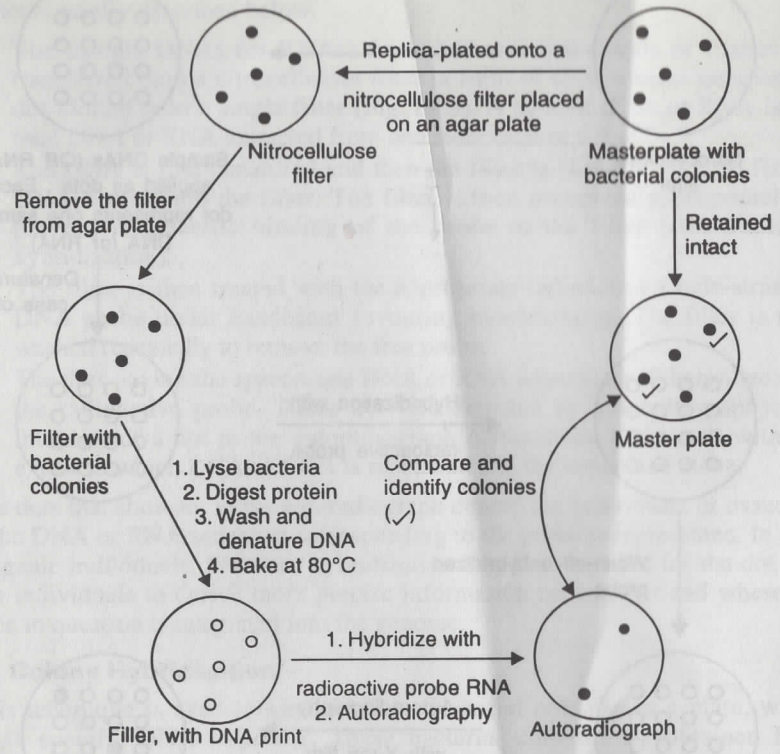


Fig. 27.6. Colony hybridization scheme is used to identify bacterial colonies containing a specific sequence of DNA, which is represented by the radioactive DNA or RNA probe used in the test.

27.4.7. Chromosome Walking

This technique is used for characterising large regions of chromosomes. Using the conventional cloning techniques to isolate fragments, one may be able to characterise about a 100 kb chromosome region per month. In addition, pulsed field gel electrophoresis may be used to isolate >100 kb DNA fragments, which may be cloned. Generally, a cosmid library is used for chromosome walking; each clone in such a library may be expected to have an insert, on an average, of 50 kb. In ***chromosome walking***, one begins with a DNA fragment that contains a known gene/genetic marker. The sequence located at one end of this fragment is used to identify a clone that has such DNA insert, which partly overlaps the first fragment. Now the other (nonoverlapping) end of this new DNA fragment is used as probe. In this way, one continues to move step-by-step toward the gene of interest located close to the known gene/genetic marker. This technique, therefore, is called ***chromosome walking*** because each clone takes the researcher one step closer to the gene of interest. A generalised procedure for chromosome walking is given below (Fig. 27.7).

1. The first step in 'walking' a chromosome consists of the isolation of a DNA fragment (fragment 1) containing a known gene or marker located

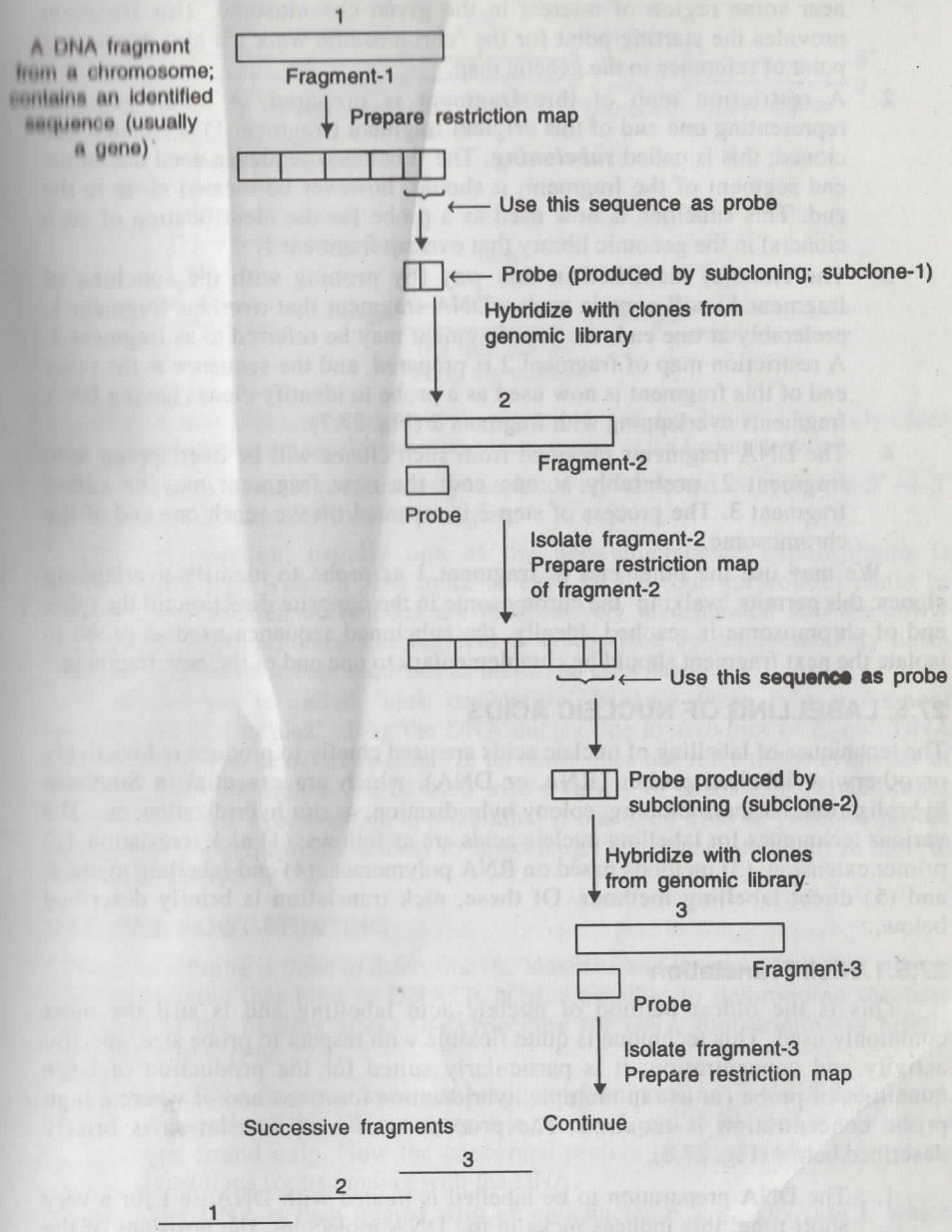


Fig. 27.7. A schematic representation of the technique of 'chromosome walking'; in this approach the sequence at one end of a fragment is used as probe to identify the next overlapping fragment and so on.

near some region of interest in the given chromosome. This fragment provides the starting point for the 'chromosome walk'; it also provides a point of reference in the genetic map.

2. A restriction map of this fragment is prepared. A small segment representing one end of this original fragment (fragment 1) is isolated and cloned; this is called ***subcloning***. The subcloned sequence need not be the end segment of the fragment; it should, however be located close to the end. This subclone is now used as a probe for the identification of such clone(s) in the genomic library that overlap fragment 1.
3. The clone(s) identified in this way (by probing with the subclone of fragment 1) will contain such a DNA fragment that overlaps fragment-1, preferably at one end; the new fragment may be referred to as fragment 2. A restriction map of fragment 2 is prepared, and the sequence at the other end of this fragment is now used as a probe to identify clones having DNA fragments overlapping with fragment 2 (Fig. 27.7).
4. The DNA fragments obtained from such clones will be overlapping with fragment 2, preferably at one end; the new fragment may be called fragment 3. The process of step 3 is repeated till we reach one end of the chromosome.

We may use the other end of fragment 1 as probe to identify overlapping clones; this permits 'walking' the chromosome in the opposite direction till the other end of chromosome is reached. Ideally, the subcloned sequence used as probe to isolate the next fragment should be complementary to one end of the new fragment.

27.5. LABELLING OF NUCLEIC ACIDS

The techniques of labelling of nucleic acids are used chiefly to produce radioactively or otherwise labelled probes (RNA or DNA), which are essential in Southern hybridization, northern blotting, colony hybridization, *in situ* hybridization, etc. The various techniques for labelling nucleic acids are as follows: (1) nick translation, (2) primer extension, (3) methods based on RNA polymerace, (4) end-labelling method, and (5) direct labelling methods. Of these, nick translation is briefly described below.

27.5.1. Nick Translation

This is the oldest method of nucleic acid labelling and is still the most commonly used. This technique is quite flexible with respect to probe size, specific activity and concentration; it is particularly suited for the production of large quantities of probe for use in multiple hybridization reactions and/or where a high probe concentration is required. The procedure of nick translation is briefly described below (Fig. 27.8).

1. The DNA preparation to be labelled is treated with DNAase I for a very short time; this induces nicks in the DNA molecules, the positions of the nicks being random. A ***'nick'*** is the point in one strand of a double-stranded DNA molecule where the phosphodiester bond has been broken; thus each nick has a free 3′—OH group.
2. The nicked DNA sample is subjected to *E. coli* DNA polymerase I, which successively adds new nucleotides to the free 3′—OH group. As the new segment of DNA is synthesized in this way, the existing copy of the strand

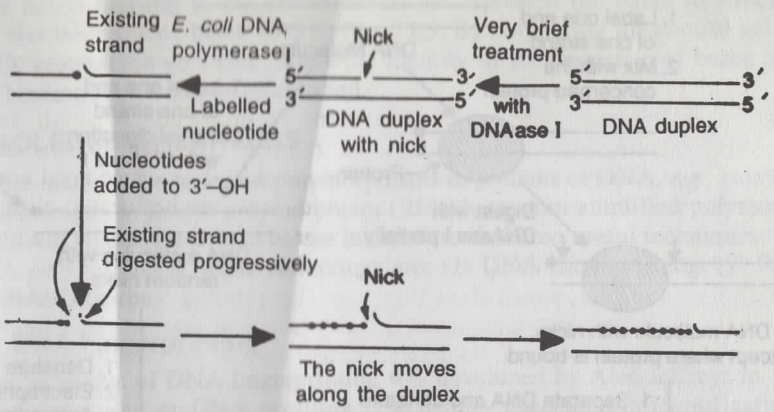


Fig. 27.8. A schematic representation of the process of 'nick translation'. The newly added nucleotides are usually radioactive (one or more of the four nucleotides).

in the duplex is progressively displaced and digested away by the 5′ → 3′ exonuclease activity of the enzyme.

In this reaction, usually one of the deoxynucleoside triphosphates is radiolabelled, *e.g.*, with ^{32}P, but more or even all the four nucleotides could be labelled. The labelled nucleotides are progressively incorporated into the newly synthesized segment; since this segment is synthesized by complementary base-pairing it has the same base sequence as the strand it replaces.

The process is called 'nick translation' because there is a movement (translation) of the 'nick' along the DNA duplex due to activities of *E. coli* DNA polymerase I. The term 'translation' in its name does not refer to the synthesis of proteins . Nick translation can be used with a variety of labels to generate probes suitable for most hybridization applications. It has been used to generate probes having high enough specific activity to detect single copy genes on Southern blots of mammalian DNA.

27.6. DNA FOOT-PRINTING

DNA foot-printing is done to determine the locations and lengths of binding sites of various proteins that bind to DNA. It is also possible to determining the base sequence of the binding sites. The procedure for DNA foot-printing is briefly described below (Fig. 27.9).

1. The DNA fragment known to bind to a given protein is isolated and cloned. A sample of this DNA fragment is prepared, and is labelled at one end of one strand only. Now the concerned protein is added under appropriate conditions for its binding with the DNA.
2. The DNA is then treated with an endonuclease, *e.g.*, DNAase I, which nicks DNA at every phosphodiester bond (or after every base); the ultimate products of digestion with such an enzyme are the four nucleoside monophosphates. An ***endonuclease*** attacks individual phosphodiester bonds within a nucleic acid. The conditions of the treatment are so chosen that a given phosphodiester bond is broken in some but not all the DNA fragments.

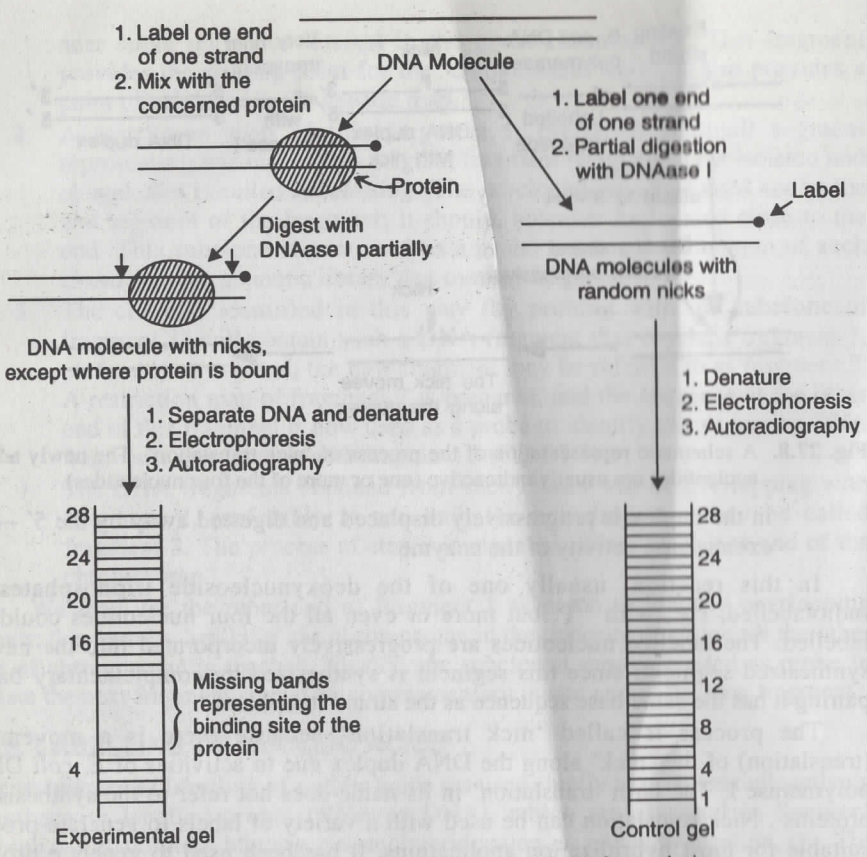


Fig. 27.9. A schematic representation of DNA foot-printing.

3. The DNA fragments are separated from the bound protein, denatured and subjected to gel electrophoresis. Each fragment that has the labelled end will appear on the X-ray film following autoradiography (***experimental gel***). The gel electrophoresis technique is sensitive enough to separate DNA fragments differing in length by a single base; the shortest fragments move the fastest. Since the endonuclease is equally likely to digest every phosphodiester bond in the DNA fragment, every successive band in the gel beginning at its lower end in Fig. 27.9 represents the successive bases from the labelled end of the concerned strand; the first band represents the labelled base itself.
   The protein bound to the DNA protects the latter from endonuclease digestion. Therefore, the bands corresponding to those bases with which the protein was bound will be absent from the gel; this becomes obvious by comparing this gel with a control gel.
4. The ***control gel*** is prepared in the same manner as described above with one critical difference: in this case, the DNA fragments are free and not bound with the protein. As a result, every phosphodiester bond is broken in one or the other molecule, and bands representing every successive base in the molecule appear as bands in the gel.

The bands missing in the experimental gel represent the bases to which the protein was bound, *e.g.*, bases 7-12 in Fig. 27.9. By comparing the control gel with the DNA sequencing gels (see, later) the identity of the sequence of bases in the protein binding region can be established.

27.7. MOLECULAR MARKERS

Molecular markers are based on polymorphisms in proteins or DNA, *e.g.*, isozymes, restriction fragment length polymorphism (RFLP), random amplified polymorphic DNAs (RAPDs), etc. Described below in brief are the three useful techniques based on DNA polymorphism; these techniques are: (1) DNA fingerprinting, (2) RFLP, and (3) RAPDs maps.

27.7.1. DNA Fingerprinting

The technique of DNA fingerprinting was developed by Alec Jeffreys in 1985. DNA fingerprinting or DNA profiling is generally used for the identification of criminals from blood stains, semen, etc. for establishing parentage in case of dispute and many other applications. The data from this approach are extremely reliable as compared to the conventional analyses of serum proteins and erythrocyte antigens and proteins. In its original form, DNA fingerprint of an individual is essentially a Southern blot of his DNA digested with an endonuclease and probed with a radioactive DNA probe. The procedure for DNA fingerprinting is briefly described below (Fig. 27.10).

1. The first step is to obtain a sample of DNA of the individual in question. The DNA is usually obtained from blood samples, but it may be isolated from several years old blood or semen stains on clothings, from vaginal swabs taken from rape victims, even from bone marrow of years old bones of murder victims, etc. The amount of DNA needed for developing fingerprints is very small (a few nanograms).
2. The DNA is disgested with a suitable restriction endonuclease, and the digest is subjected to gel electrophoresis (see, Southern hybridization).
3. The DNA is denatured and then transferred from the gel onto a nitrocellulose filter membrane, where it is fixed by baking the filter at 80°C.
4. The DNA bands on the filter are hybridized with the appropriate radioactive probe (single-stranded DNA probe); the free probe is washed off and the bands to which the probe has hybridized are detected through autoradiography.

The probes used for DNA fingerprinting are usually prepared from mini-satellite DNA. The mini-satellite DNA is highly variable mainly due to variations in the number of tandem repeats of a short core sequence. These probes hybridize, under conditions of low stringency, to a number of polymorphic loci represented by the mini-satellite DNA. The regions of minisatellite may occur dispersed throughout the human genome or may be clustered on to a single chromosome. A simple universal probe, a tanden repeat of GATA, has been developed from the sex chromosomes of banded karait; this probe is reportedly very useful in DNA fingerprinting of not only man but of many other organisms, including plants. More recent strategies of DNA fingerprinting use PCR to amplify minisatellite/ microsatellite sequences.

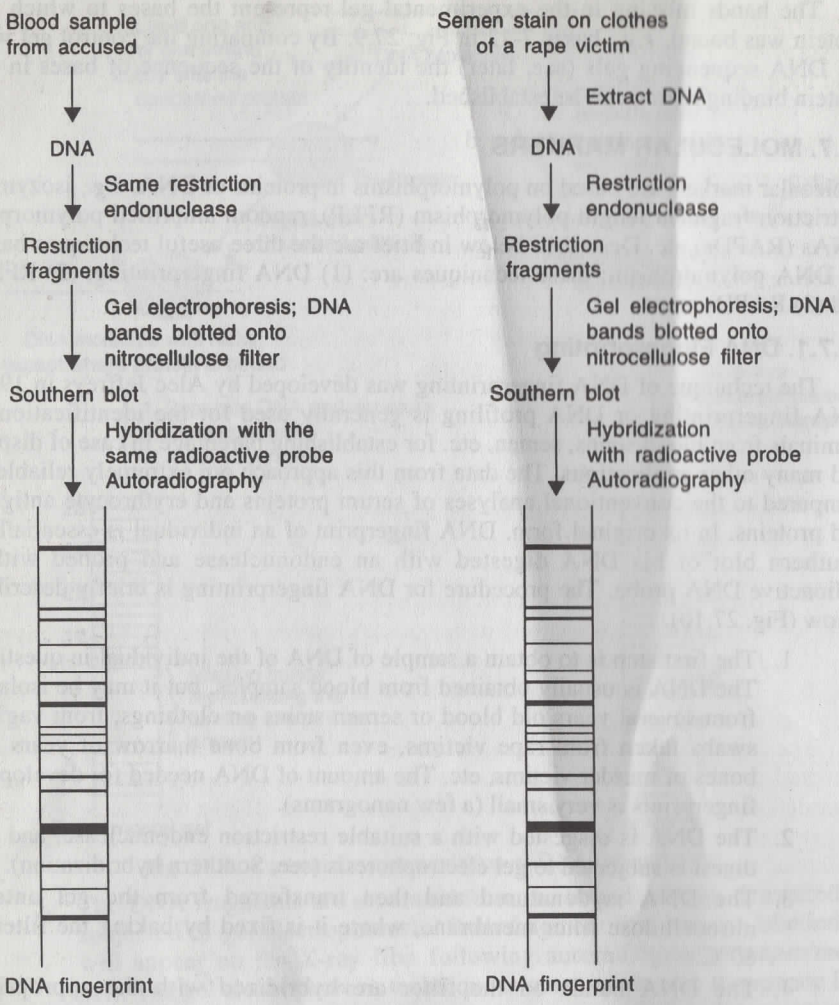


Fig. 27.10. A schematic representation of DNA fingerprinting of a person accused of rape, and of the semen stain recovered from the clothes of the rape victim. The fingerprints match perfectly; the probability of this being due to chance is extremely low ordinarily, a control, *viz.*, DNA fingerprint of an unrelated person, is also included in the test.

The bands appearing in the DNA fingerprint of the suspected criminal(s) are compared with those obtained from the test DNA recovered from the evidence for crime and believed to originate from the criminal, *e.g.*, semen stains or sperm nuclei in vaginal swabs from rape victims; etc.; there should be a perfect match in case the suspect had been involved. The fingerprints of the victim and of an unrelated individual are always included as conrols for comparison.

In cases of disputed parentage, the DNA fingerprints of the child, the mother and the suspected father are compared. As a rule, the bands present in a child's

fingerprint must be accounted for by those in the mother's and the father's fingerprints. Therefore, the child's fingerprint is compared with that of the mother and their common bands are marked. The remaining bands in the child's fingerprint must be present in that of the child's father. DNA fingerprint of a sib (brother or sister of the child) is always included for comparison as control.

27.7.2. RFLP

RFLPs (restriction fragment length polymorphisms) denote that a single restriction enzyme produces fragments of different lengths from the same stretch of genomic DNA of different strains or related species. These differences are detected as bands differing in mobilities on the Southern blots of these strains/species. A generalized procedure for studying RFLPs is as follows.

1. Large molecular weight genomic DNAs are isolated from several strains or related species.
2. Each of these DNAs is then digested with the same selected restriction enzyme.
3. The fragments in these digests are separated by gel electrophoresis, each digest being assigned a separate lane in the same gel.
4. The resulting gel lanes are denatured and transferred onto a suitable solid support, *e.g.*, nitrocellulose filter membrane, and the DNA fragments are fixed to it firmly (Southern blotting).
5. The filter is hybridized with the appropriate radioactive single-stranded DNA probe (Southern hybridization), and the free probes are washed off.
6. The bands with which the selected probe has hybridized are detected by autoradiography.

The lanes of the different strains/related species are compared, and RFLPs are detected as differential movement of a band on their gel lanes; each such band is regarded as a single ***RFLP locus.*** It may be noted that RFLPs are the same as DNA fingerprints described above, and are often used for the unequivocal identification of different varieties of a crop species.

Thus each RFLP locus is definable only by the combination of a specific restriction enzyme with a specific DNA probe. The DNA probes may be obtained from the following: (*i*) genomic libraries, (*ii*) cDNA libraries (these may be random or specific), or (*iii*) chromosome-specific libraries obtained in various ways. Single-copy sequences (representing, most likely, structural genes) are the best probes, but low-copy and even multiple copy sequences are also used. Generally, probes prepared from the same species (*homologous probes*) are used, but those from other species (*heterologous probes*) may also be employed.

The pattern of RFLPs will depend mainly on the following: (*i*) differences in the DNAs of the selected strains/species, (*ii*) the restriction enzyme used, and (*iii*) the DNA probe employed in Southern hybridization. Detectable RFLPs are generated due to the following changes in the DNA: (*i*) changes in the base sequences of the recognition sites of the concerned restriction enzymes, and (*ii*) relatively large deletions/additions in the genomic DNA between the two restriction sites.

The linkage among different RFLP loci is determined by studying suitably developed populations and, using these data, a map is developed; this is called ***RFLP map***. The RFLP map can be integrated with genetic and cytological maps prepared by conventional techniques. The RFLP maps can also be assigned to individual chromosomes. RFLP maps are useful for the following: (*i*) identification

and isolation of any gene linked with known RFLP loci, (*ii*) unequivocal identification of strains/varieties, (*iii*) linkage mapping of quantitative trait loci (QTLs), (*iv*) indirect selection for QTLs tightly linked with RFLP loci, and (*v*) for several other purposes.

27.7.3. RAPDs

RAPDs (Randomly amplified polymorphic DNAs) are produced by using a PCR (polymerase chain reaction) equipment or a thermal cycler. The procedure, in simple terms, for obtaining RAPDs is as follows.

1. The genomic DNA of a selected strain/variety/species is isolated in a high molecular weight condition.
2. To this DNA is added an excess of a selected oligonucleotide (a short polynucleotide; usually, 10 bases long) that serves as the primer. The oligonucleotide is usually obtained by *in vitro* DNA synthesis for which often a sophisticated equipment (called gene machine or obligonucleotide synthesizer) is used.
3. This mixture is subjected to repeated cycles of DNA denaturation—renaturation—DNA replication in a PCR equipment. During renaturation, following denaturation, the oligonucleotide will pair with the homologous sequence present at different locations in the genomic DNA; therefore, DNA replication will extend the oligonucleotide and copy the sequence continuous with the sequence with which the selected oligonucleotide had paired. The repeated cycles of denaturation—renaturation—DNA replication will, therefore, amplify this sequence of the genomic DNA. Amplification will take place only of those regions of the genome that have the sequence complementary to the random primer at both their 5′-ends (Fig. 27.11).
4. After several cycles of amplification, the DNA is subjected to gel electrophoresis. The amplified DNA will form a distinct band, which is usually detected by ethidium bromide staining and visible fluorescence under UV light.

It may be pointed out that RAPD approach does not utilize restriction enzymes (which are very costly) and probes of any kind; therefore, they save both cost and effort in comparison to RFLPs. They have the same use as RFLPs, and are gaining ground over the latter.

The RAPD technique is based on PCR, which uses a DNA polymerase that is resistant to the high temperature needed for DNA denaturation; this permits the enzyme to be used repeatedly during the amplification process. The oligonucleotide used may represent the base sequence of a short segment of the genomic DNA that is already known; variation in the sequence of this segment can now be detected by the RAPD approach. However, when this information is either unavailable or is not desirable for use, oligonucleotides having random base sequence can be used.

Those strains that have a segment of DNA both the ends of which are homologous to the oligonucleotide used, a distinct band will be seen on the agarose gel. However, this band will be absent in the strains in which one or both the ends of this sequence of genomic DNA is either modified or deleted. Thus RAPD bands are either 'present' or 'absent' in different strains. Further, the F_1 from the cross 'present' × 'absent' strains for an RAPD band will always show the band; thus

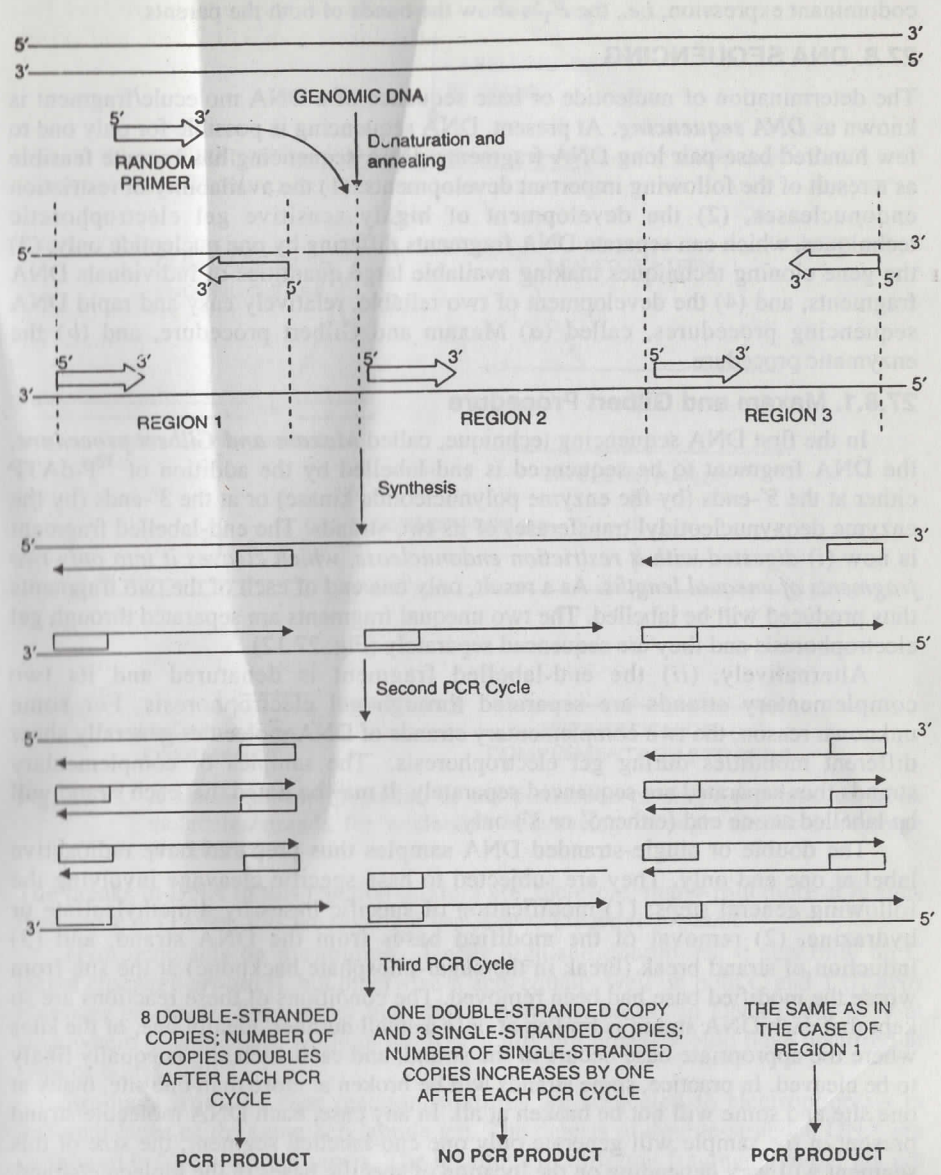


Fig. 27.11. Schematic representation of events during RAPD detection. A region of the genome will be amplified only when it is flanked by a sequence at its both 5′-ends that is complementary to the random primer used, *e.g.*, regions 1 and 3. In case, the complementary sequence occurs at only one side of a region, *e.g.*, region 2, amplification cannot take place. The length of amplified region depends on many factors, and may go upto 10 kb.

RAPD bands have dominant expression, while RFLP bands, in contrast, show codominant expression, *i.e.*, the F_1's show the bands of both the parents.

27.8. DNA SEQUENCING

The determination of nucleotide or base sequence of a DNA molecule/fragment is known as ***DNA sequencing.*** At present, DNA sequencing is possible for only one to few hundred base-pair long DNA fragments. DNA sequencing has become feasible as a result of the following important developments: (1) the availability of restriction endonucleases, (2) the development of highly sensitive gel electrophoretic techniques, which can separate DNA fragments differing by one nucleotide only, (3) the gene cloning techniques making available large quantities of individuals DNA fragments, and (4) the development of two reliable, relatively easy and rapid DNA sequencing procedures, called (*a*) Maxam and Gilbert procedure, and (*b*) the enzymatic procedure.

27.8.1. Maxam and Gilbert Procedure

In the first DNA sequencing technique, called ***Maxam and Gilbert procedure***, the DNA fragment to be sequenced is end-labelled by the addition of ^{32}P-dATP either at the 5′-ends (by the enzyme polynucleotide kinase) or at the 3′-ends (by the enzyme deoxynucleotidyl transferase) of its two strands. The end-labelled fragment is now (*i*) *digested with a restriction endonuclease, which cleaves it into only two fragments of unequal lengths.* As a result, only one end of each of the two fragments thus produced will be labelled. The two unequal fragments are separated through gel electrophoresis and they are sequenced separately (Fig. 27.12).

Alternatively, (*ii*) the end-labelled fragment is denatured and its two complementary strands are separated through gel electrophoresis. For some unknown reason, the two complementary strands of DNA molecules generally show different mobilities during gel electrophoresis. The samples of complementary strands thus separated are sequenced separately. It may be noted that each strand will be labelled at one end (either 5′ or 3′) only.

The double or single-stranded DNA samples thus prepared have radioactive label at one end only. They are subjected to base specific cleavage involving the following general steps: (1) modification of specific bases by dimethylsulfate or hydrazine, (2) removal of the modified bases from the DNA strand, and (3) induction of strand break (break in the sugar-phosphate backbone) at the site from where the modified base had been removed. The conditions of these reactions are so kept that each DNA strand is broken at only a small number, ideally one, of the sites where the appropriate base occurs in the strand, and each such site is equally likely to be cleaved. In practice, some strands will be broken at more than one site, many at one site and some will not be broken at all. In any case, each DNA molecule/strand present in the sample will generate only one end-labelled segment; the size of this segment will vary depending on the location of specific bases in the molecule/strand.

Dimethylsulfate adds a methyl residue to N7 position of G and to N3 position of A, the methylation of G being ~5 times more frequent than that of A. When DNA treated with dimethylsulfate is heated, the methylated bases are lost from DNA strands, and the strands break at the sites of base loss. The breaks are ~5 times more frequent at sites of G than those at sites of A; this constitutes the ***first cleavage product of DNA.*** But when dimethylsulfate-treated DNA is subjected to acid treatment, methylated A is lost several times more frequently than is methylated G.

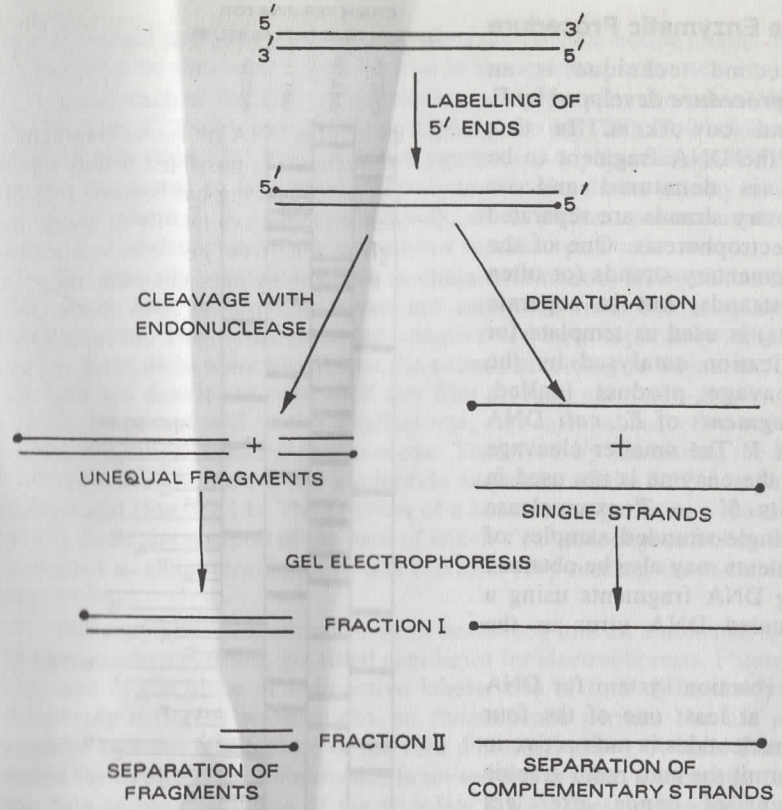


Fig. 27.12. A schematic representation of the production of single-end labelled DNA molecules/strands for nucleotide sequencing through chemical cleavage procedure.

The breaks in DNA in this reaction mixture are, therefore, many-fold more at A sites than those at G sites; this we regard as the ***second cleavage product.***

Hydrazine acts on pyrimidines. Under high salt conditions, it acts only on C (***third cleavage product***), but otherwise it acts equally well on both C and T (***fourth cleavage product***). Thus under high salt conditions, breaks in DNA occur only at sites where C is located, while in the fourth mixture breaks would have occurred at sites of both C and T.

Products from the four reaction mixtures are subjected to electrophoresis in different lanes of the same acrylamide gel. The digests of double-stranded DNA are denatured before electrophoresis. The locations of different fragments in the gel are determined by autoradiography, which detects only such fragments that contain the labelled end of DNA. The bands appearing in the 4 lanes of the gel will represent the locations of the following bases: (1) G (darker bands, lighter bands will be A sites; ***cleavage product one***), (2) A (darker bands; ***cleavage product two***), (3) C (***cleavage product three***), and (4) C + T (***cleavage product four***). Base sequence of the DNA is determined by sequentially reading together the bands developed in the four lanes of the gel; each successive band represents a DNA strand differing from the previous one by just one base.

27.8.2. The Enzymatic Procedure

The second technique is an ***enzymatic procedure*** developed by F. Sanger and coworkers. In this technique, the DNA fragment to be sequenced is denatured and its complementary strands are separated through electrophoresis. One of the two complementary strands (or often both the strands, but in separate experiments) is used as template for DNA replication catalysed by the larger cleavage product (called, ***Klenow fragment***) of *E. coli* DNA polymerase I. The smaller cleavage product of the enzyme is not used in view of its 5′ → 3′ exonuclease activity. Single-stranded samples of DNA fragments may also be obtained by cloning DNA fragments using a single-stranded DNA virus as the vector.

In the reaction system for DNA replication, at least one of the four deoxyribonucleotides is radioactive in order to permit the auto radio graphic development of bands after gel electrophoresis. A small primer sequence with a free 3′—OH group must be provided with the template strand for DNA replication to proceed, since a free 3′—OH group is absolutely essential for DNA polymerase I to catalyse DNA replication.

Four different reaction mixtures are prepared for the replication of each DNA fragment to be sequenced. In one of the reaction systems, 2′, 3′-dideoxycytidine triphosphate (ddCTP) is added in a concentration about 1/100th of the normal deoxycytidine triphosphate present in the system. ddCTP acts as a terminator of the polynucleotide chain being newly synthesized on the template strand. Chain termination by ddCTP is achieved due to the fact that it (and the other 2′, 3′-dideoxynucleotides) does not have a free 3′–OH group as a result of which further nucleotides cannot be added to the new chain. At the concentration used here,

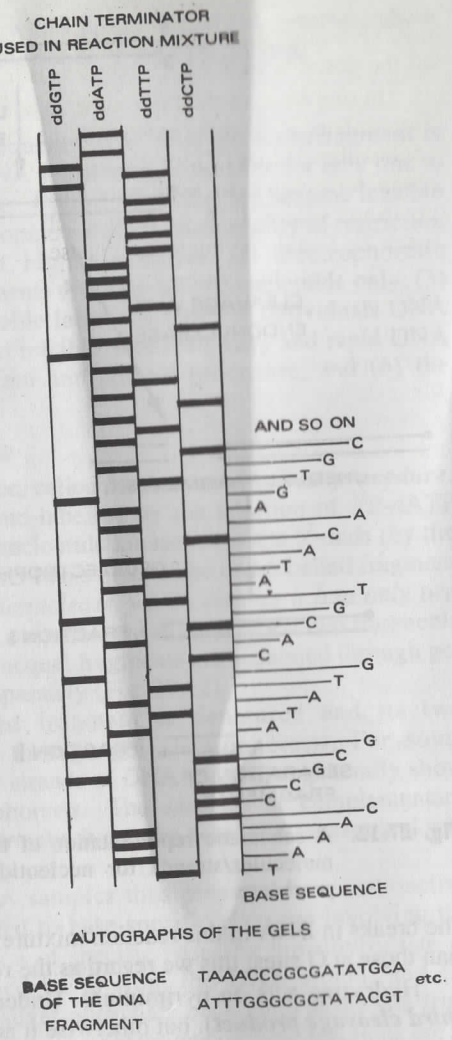


Fig. 27.13. Diagrammatic representation of the bands obtained in a DNA sequencing gel. The single-stranded fragment was used as template for DNA synthesis and chains were terminated using either ddGTP, ddATP, ddTTP or ddCTP. The shortest fragments are the bottom-most; only a small segment of the gel is shown.

ddCTP would terminate the newly synthesized polynucleotide chains at any one of all the possible sites where cytosine is to be incorporated in the new chain.

In one each of the three other reaction mixtures using the same DNA fragment as template, 2′, 3′-dideoxy-thymidine triphosphate (ddTTP), 2′, 3′-dideoxyadenosine triphosphate (ddATP), or 2′, 3′-dideoxyguanosine triphosphate (ddGTP) is used as chain terminator to terminate the polynucleotide chains at any one of all the positions where T, A or G, respectively, are to be incorporated in the new chain (each 2′, 3′-dideoxynucleotide is used in a separate reaction mixture).

The partially synthesized (due to chain termination) DNA chains from each of the above four reaction mixtures are separated from the template strand by denaturation. The single stranded samples are now separately subjected to gel electrophoresis in order to separate the strands according to their size. The bands in the gels are developed onto an X-ray film through autoradio-graphy. The fastest moving fragment will be the smallest one, and each subsequent band will be one nucleotide longer than the previous one. Therefore, by comparing the bands of the four gels thus obtained, the nucleotide sequence of the DNA fragment can be determined (Fig. 27.13). The position of a band in the gel from a reaction mixtures will indicate the position of the base of which 2′, 3′-dideoxynucleotide tri-phosphate was used as chain terminator in that mixture. (Try reading the remaining bases in Fig. 27.13).

The enzymatic procedure has been developed into an automated system using, in most modern systems, gel filled capillaries for electrophoresis. Fluorescent labels are used in the place of radioactive labels. Use of four different dyes that give detectably different wavelengths on fluorescence permits the use of only one reaction mixture in the place of the four described above. A laser beam is used to excite the dyes and the fluorescence is sensed and fed into a computer that interprets the data as base sequence of the template DNA. These automated systems were developed mainly for the needs of the human genome project, and enabled the project to be completed before its initially projected date.

QUESTIONS

1. List the common molecular techniques used in genetic studies. Describe briefly the principle of DNA renaturation studies and the information derived from them.
2. A DNA sequence has been introduced through genetic engineering into a population of bacterial cells and subsequently into several plants. Briefly describe the techniques used for the following.
   (*i*) Identification of the bacterial colonies in which this sequence is present.
   (*ii*) Identification of the plants into which this sequence is present.
   (*iii*) Identification of the DNA fragment of plant genome into which this sequence is integrated.
3. Briefly describe the northern and western blotting techniques and discuss their applications.
4. Describe briefly the techniques of DNA fingerprinting, and of RFLP and RAPD detection, and discuss their applications.
5. Write short notes on the following:
   (*i*) heteroduplex mapping, (*ii*) restriction mapping,
   (*iii*) nick translation, (*iv*) chromosome walking,
   (*v*) DNA sequencing, (*vi*) DNA footprinting,
   (*vii*) RAPDs.

28

Gene Action I. Genes, Biochemical Reactions and Proteins (Biochemical Genetics)

28.1. INTRODUCTION

Genes produce specific phenotypic effects through a series of biochemical reactions. In such a reaction chain, the product of the previous reaction serves as the substrate for the next reaction. Each reaction in the chain is catalysed by a specific enzyme. Each enzyme, in turn, is encoded by a specific gene. In addition, some genes produce nonenzyme or structural proteins (Fig. 28.1). Thus the function of genes is to produce specific enzyme or nonenzyme proteins. The enzymes, in turn, catalyse specific biochemical reactions, which produce phenotypic effects.

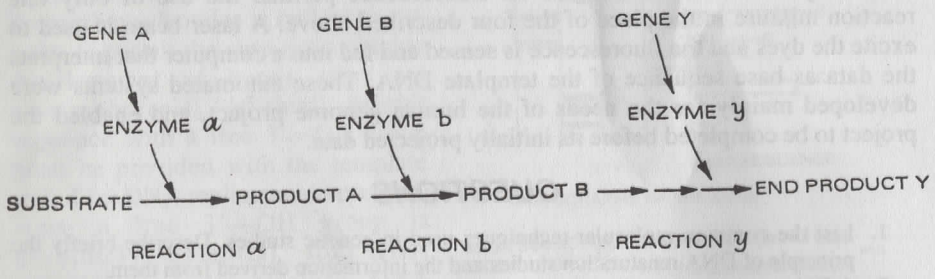


Fig. 28.1. A generalized scheme for the biosynthesis of Y, the end-product of a series of biochemical reactions. Each reaction in the biosynthetic chain is catalyzed by a specific enzyme, which, in turn, is produced by a specific gene.

28.2. GENETIC CONTROL OF BIOCHEMICAL REACTIONS

That genes control biochemical reactions was first proposed by Sir A.E. Garrod, an English physician, in 1909. He concluded that in alkaptonuria patients, some enzyme responsible for alkapton metabolism was either inactive or missing. As a consequence, alkaptonuria patients excrete alkapton in their urine, which turns black on exposure to air. Alkapton or homogentisic acid is produced during breakdown of phenylalanine and tyrosine. He reached a similar conclusion regarding albinism, a recessive trait (*cc*) in which the pigment melanin is absent from hairs, retina and skin cells. Subsequently, during 1930's, a group of English workers began studies on the biochemistry of flower pigments. They demonstrated that there was a precise correlation between specific genes and biochemical events.

28.2.1. *Drosophila* Eye Transplants

Further evidence that genes control biochemical reactions came in 1937 from the studies on *Drosophila* eye pigments by Beadle and Ephrussi. They transplanted embryonic eye tissue (***imaginal discs***) from 26 different eye colour mutants on to the abdomen of wild type larvae. The transplanted imaginal discs developed into an additional eye. Through an analysis of the colour of the additional eyes in suitably constructed transplants, they concluded that the block in pigment production in vermilion was located earlier in the reaction chain than that in the cinnabar (Fig. 28.2). It is now known that mutation in gene *vermilion* (*v*) blocks the conversion of tryptophan to N-formylkynurenine, while *cinnabar* (*cn*) controls the formation of 3-hydroxykynurenine from kynurenin. (Kynurenine is itself produced from N-formyl-kynurenine.)

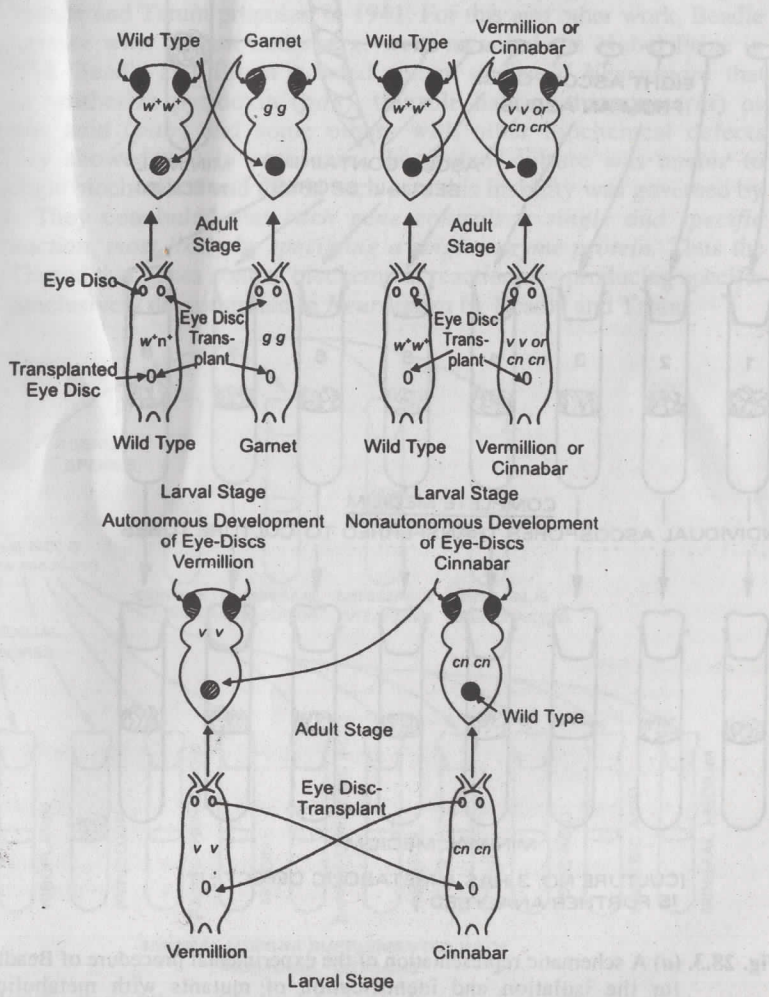


Fig. 28.2. Autononomous and nonautonomous development of eye colour in *Drosophila* eye disc transplants into the larvae of different genotypes.

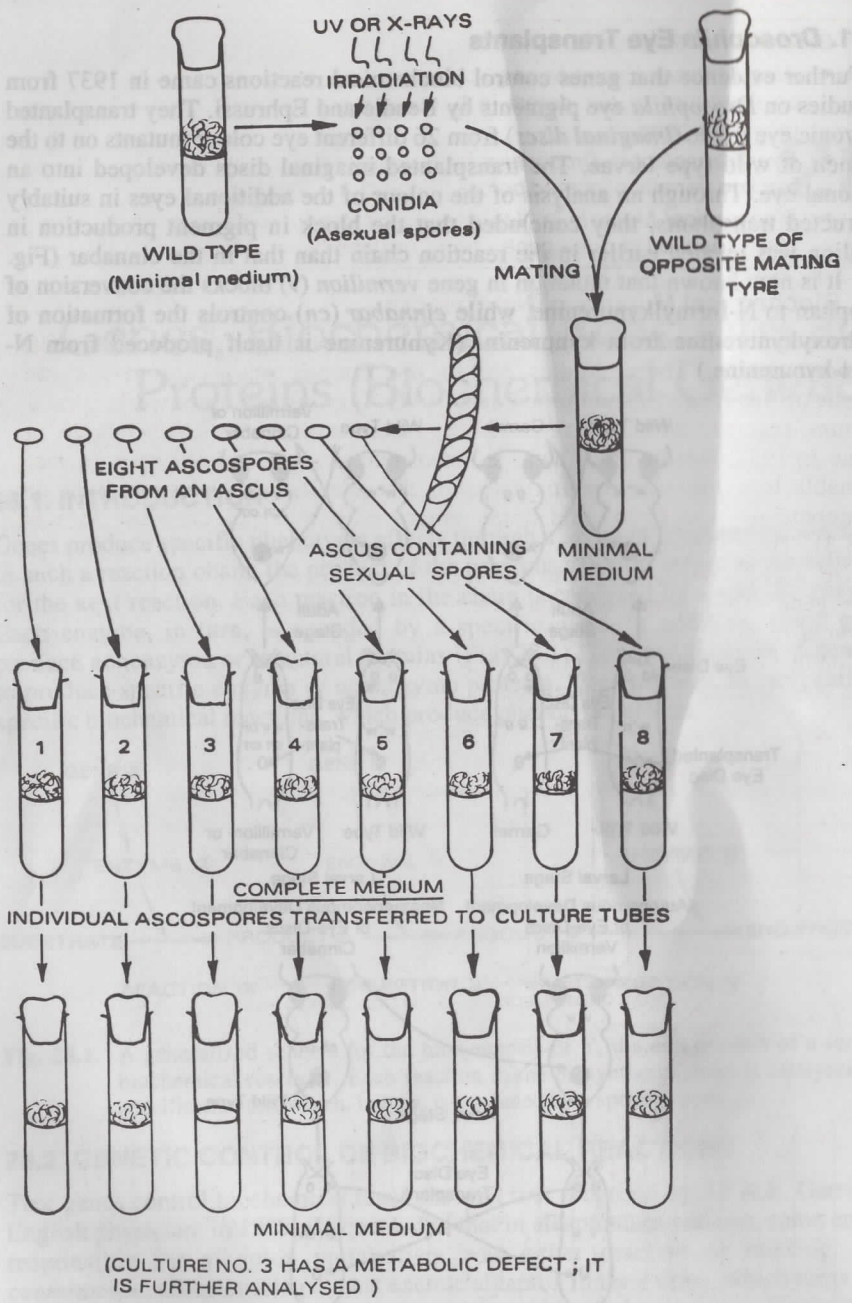


Fig. 28.3. (*a*) A schematic representation of the experimental procedure of Beadle and Tatum for the isolation and identification of mutants with metabolic defects in *Neurospora*. In this case, culture No. 3 has a metabolic defect; it is unable to synthesize some biochemical essential for its growth.

Biochemical Reactions

- *Alkaptonurea*: Excretion of alkapton in urine, which turns black on exposure to air; defect in tyrosine metabolism; due to a recessive gene.
- *Eye Transplant in Drosophila*: by Beadle and Ephrussi (1937); block in vermilion located before the block in cinnabar.
- *Biochemical Mutants in Neurosporn*: by Beadle and Tatum in 1941; resulted in one - gene - one - enzyme hypothesis.

28.2.2. Biochemical Mutants of Neurospora

Subsequently, Beadle investigated nutritional mutants in red bread mould, *Neurospora crassa*. These studies led to the famous one-gene-one-enzyme hypothesis of Beadle and Tatum proposed in 1941. For this and other work, Beadle and Tatum, together with Joshua Lederberg, were awarded the Nobel Prize in Medicine in 1958. Beadle and Tatum isolated mutant strains of *Neurospora* that were unable to synthesize pyridoxin (*pdx*), thiazole half of thiamine (*thi*) or *p*-amionobenzoic acid (*pab*) and some others with other biochemcal defects (Fig. 28.3). They showed that in each case, the mutant culture was unable to synthesize a single biochemical and that in each case this inability was governed by a single gene. They concluded that *each gene controls a single and specific biochemical reaction, most likely by specifying a single enzyme protein.* Thus the suggestion by Garrod that genes control biochemical reactions by producing specific enzymes was conclusively demonstrated in *Neurospora* by Beadle and Tatum.

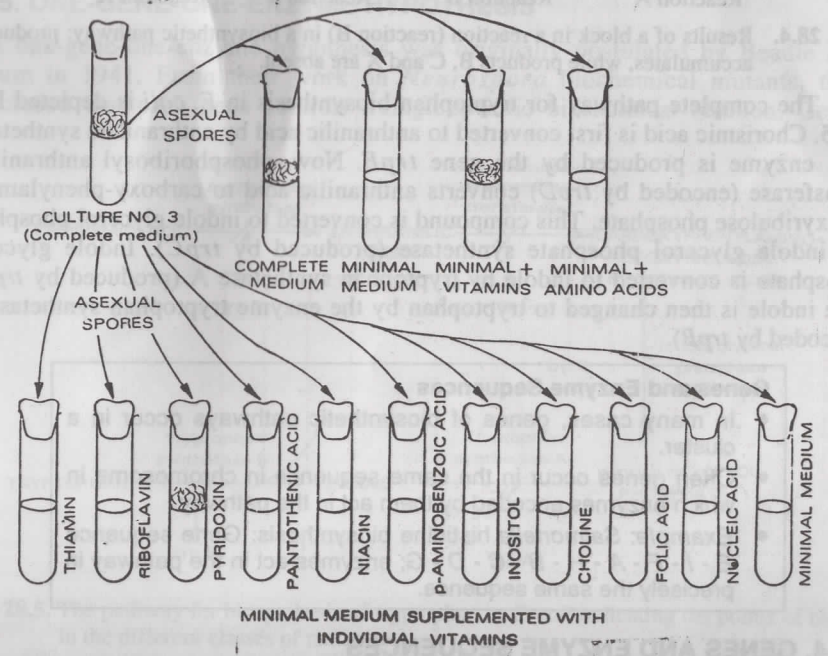


Fig. 28.3(*b*) Identification of the metabolic defect in culture No.3 identified as biochemical mutant (fig. 28.3 (*a*). The culture 3 is unable to synthesize pyridoxine.

Biosynthetic Pathway

- The precise sequence of reactions that lead to the biosynthesis of a biochemical.
- Determined by using biochemical mutants.
- Product of preceding reaction in the pathway serves as substrate for the next reaction.
- *Example:* Tryptophan biosynthesis.

28.3. BIOSYNTHETIC PATHWAYS

As a rule, the reactions involved in biosynthesis of a compound occur in a precise sequence so that product of the previous reaction serves as substrate for the next reaction. Such a sequence of reactions is known as a ***biosynthetic*** or ***metabolic pathway.*** If a particular reaction is blocked in a biosynthetic pathway, the substrate for that reaction will accumulate, while the products of that and all subsequent reactions will be absent. For example, if reaction *B* is blocked (Fig. 28.4), product A will accumulate, while products *B*, *C* and *X* will be absent. A mutant that has defect in a biochemical pathway is called ***biochemical mutant***; such mutants have been extensively used in determining the reactions of various metabolic pathways.

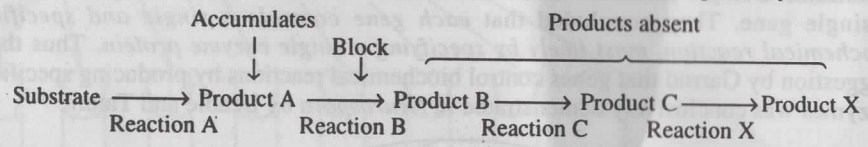


Fig. 28.4. Results of a block in a reaction (reaction B) in a biosynthetic pathway: product A accumulates, while products B, C and X are absent.

The complete pathway for tryptophan biosynthesis in *E. coli* is depicted Fig. 28.5. Chorismic acid is first converted to anthranilic acid by anthranilate synthetase; this enzyme is produced by the gene *trpE*. Now, phosphoribosyl anthranilate transferase (encoded by *trpD*) converts anthranilic acid to carboxy-phenylamino deoxyribulose phosphate. This compound is converted to indole glycerol phosphate by indole glycerol phosphate synthetase (produced by *trpC*). Indole glycerol phosphate is converted to indole by tryptophan synthetase A (produced by *trpA*). The indole is then changed to tryptophan by the enzyme tryptophan synthetase B (encoded by *trpB*).

Genes and Enzyme Sequences

- In many cases, genes of biosynthetic pathways occur in a cluster.
- Often genes occur in the same sequence in chromosome in which enzymes encoded by them act in the pathway.
- *Example*: *Salmonella* histidine biosynthesis: Gene sequence *E - I - F - A - H - B - C - D - G*; enzymes act in the pathway in precisely the same sequence.

28.4. GENES AND ENZYME SEQUENCES

In bacteria and viruses, genes governing the enzymes of a metabolic pathway are, in many cases, clustered together. Often the sequence in which these genes are

arranged in the chromosomes closely corresponds to the sequence in which the enzymes produced by them participate in the concerned metabolic pathways. For example in *Salmounella,* histidine biosynthesis involves nine genes located in a cluster; the order of these genes is *E - I - F - A - H - B - C - D - G*. This is precisely the sequence in which the enzymes produced by them participate in histidine biosynthesis, except that the last gene *G*, controls the first enzyme of the pathway. A similar close correspondence between gene arrangement and sequence of enzyme action has been shown for tryptophon, proline, threonine and isolencine-valine biosyntheses in *Salmonella,* tryptophan biosynthesis and arabinose fermentation in *E. coli,* etc. However, such a correlation is ordinarily lacking in eukaryotes.

One-Gene-One-Enzyme Hypothesis

- Based on studies on auxotrophic mutants of *Neurospora.*
- *Auxtrophic mutant*: The mutant requires a specific biochemical for normal growth.
- Each mutant is unable to synthesize a single biochemical.
- Each mutant has mutated for a single gene.
- *Conclusion*: Genes govern biodremical reactions by controlling enzyme production.
- *Hypothesis*: Each gene specifies production of one enzyme.
- *One gene - one - polypeptide hypothesis*: Due to Ingram (1957); each gene specifies one polypeptide; *universally accepted.*

28.5. ONE-GENE-ONE-ENZYME HYPOTHESIS

The one-gene-one-enzyme hypothesis was originally postulated by Beadle and Tatum in 1941. From their work on *Neurospora* biochemical mutants, they concluded that each gene controls a single specific biochemical reaction. Genes

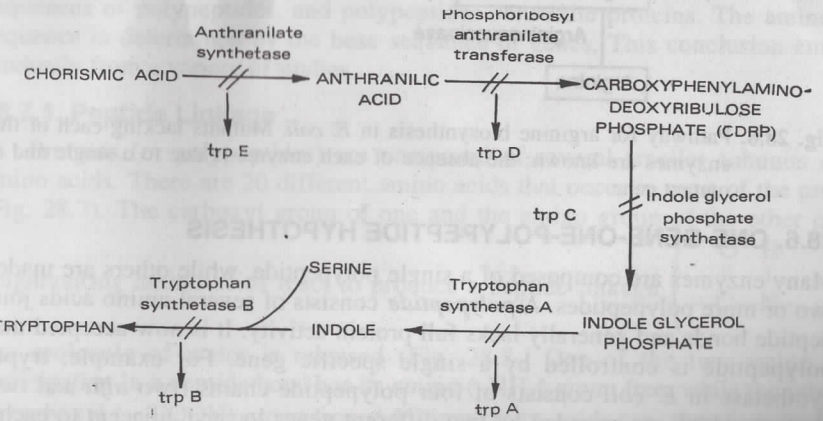


Fig. 28.5. The pathway for biosynthesis of tryptophan in *E. coli* indicating the points of block in the different classes of mutants.

control chemical reactions by controlling the production of specific enzym catalyze these reactions. Thus ***one-gene-one-enzyme hypothesis*** simply st

each gene specifies a single specific enzyme. One of the first clear-cut proofs for this hypothesis came from arginine biosynthesis pathway of *E. coli.* This pathway consists of 8 biochemical reactions, each of which is catalyzed by a specific enzyme. In turn, the production of each enzyme is controlled by a specific single gene (Fig. 28.6).

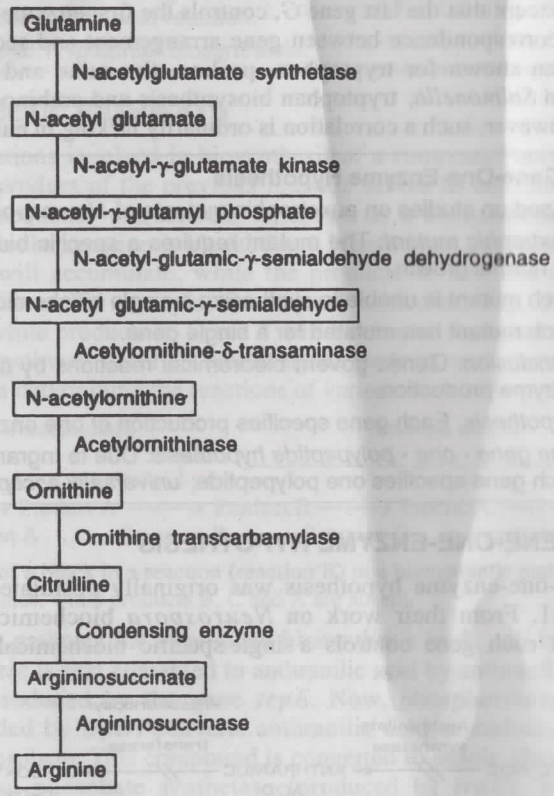


Fig. 28.6. Pathway for arginine biosynthesis in *E. coli.* Mutants lacking each of the eight enzymes are known; the absence of each enzyme is due to a single and distinct mutant gene.

28.6. ONE-GENE-ONE-POLYPEPTIDE HYPOTHESIS

Many enzymes are composed of a single polypeptide, while others are made up of two or more polypeptides. A ***polypeptide*** consists of several amino acids joined by peptide bonds and generally lacks full protein activity. It is now accepted that each polypeptide is controlled by a single specific gene. For example, tryptophan synthetase in *E. coli* consists of four polypeptide chains (two *alfa* and two *beta* chains), which are encoded by two different genes located adjacent to each other. Similarly, a haemoglobin molecule has four polypeptide chains; two chains are copies of α polypeptide and two are copies of β polypeptide. A haemoglobin molecule, therefore, can be written as $\alpha_2 \beta_2$. The α and β chains are specified by two different genes located in two different chromosomes. *Thus the one-gene-one-enzyme hypothesis is valid for only those enzymes that are composed of a single*

polypeptide. In addition, many genes specify nonenzyme protein molecules, *e.g.,* haemoglobin, myoglobin, structural proteins, etc. Therefore, the ***one-gene-one-polypeptide hypothesis*** *proposed by Ingram in 1957 is more appropriate and accounts for monomeric as well as dimeric, etc. enzymes, and for nonenzyme proteins. According to this hypothesis, a gene specifies a single polypeptide chain.*

It may be pointed out that many genes do not produce polypeptides; they only encode RNA molecules that perform various functions. Examples of such genes are tRNA genes, genes encoding ribosomal RNA (rRNA) and those specifying snRNAs (small nuclear RNAs). *Thus one-gene one-enzyme or poly peptide hypothesis would be applicable to only those genes that encode polypeptide sequences.*

Those genes that encode a diffusable product, *i.e.,* either a protein or an RNA product, are called ***structural genes***. In contrast, many DNA sequences serve as binding sites for proteins that regulate normal genetic functions. Many of these protein binding sites were earlier known as genes, *e.g.,* promoter and operator genes.

Protein Structure

- *Peptide Linkage*: –C(=O)–N(H)– bond; links amino acids to form a polypeptide.
- *Primary Structure*: Amino acid sequence.
- *Secondary Structure*: Coiling pattern, *e.g.,* α – helix.
- *Tertiary Structure*: Folding pattern
- *Quaternary Structure:* Association of two or more similar or dissimilar polypeptides to yield a functional molecule.

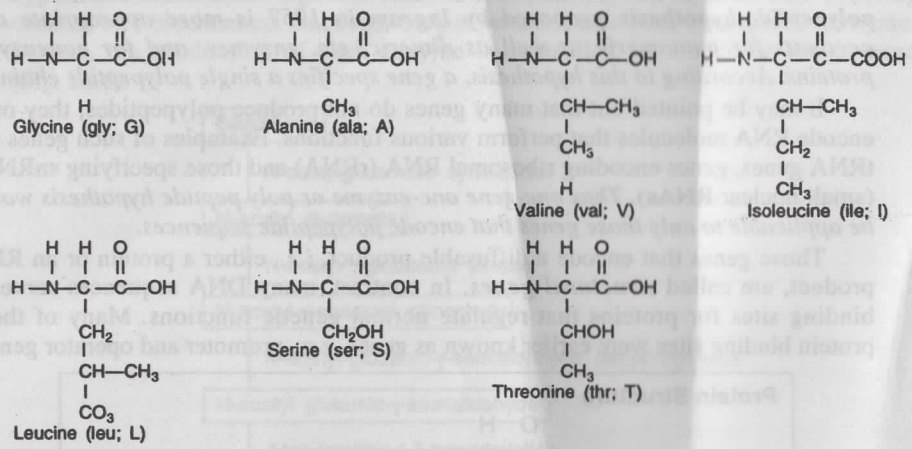
28.7. GENETIC CONTROL OF PROTEIN STRUCTURE

It has been stated that genes encode proteins. In fact, genes specify the amino acid sequences of polypeptides, and polypeptides constitute proteins. The amino acid sequence is determined by the base sequence of genes. This conclusion emerged gradually from a variety of studies.

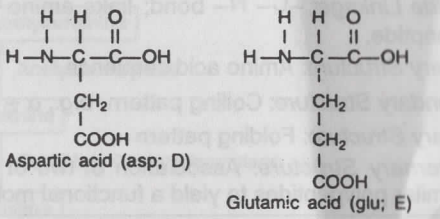
28.7.1. Peptide Linkage

Proteins (or polypeptides) are composed of several smaller subunits called amino acids. There are 20 different amino acids that occur in most of the proteins (Fig. 28.7). The carboxyl group of one and the amino group of the other of two neighbouring amino acids react to produce a peptide linkage $\left(\begin{matrix} O & H \\ \| & | \\ -C- & N- \end{matrix} \right)$ and one molecule of water is released (Fig. 28.8.) One of the two amino acids participating in a peptide bond has its amino ($-NH_2$) group free, while the other has its carboxyl (—COOH) group free. Additional mino acids can, therefore, be linked to these amino acids to form a polypeptide chain. One end of any polypeptide chain will, as a rule, have an amino group free, while the other will has have a —COOH group free; these ends are called N- and C- end, respectively, of the polypeptide. The peptide linkage of amino acids leading to polypeptide formation was elucidated by Emil Fisher.

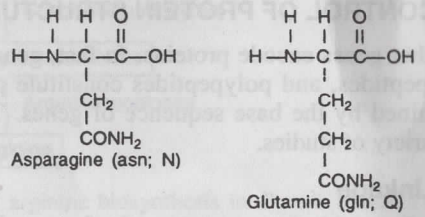
(1) Containing one amino (—NH_2) and one carboxyl (—COOH) group



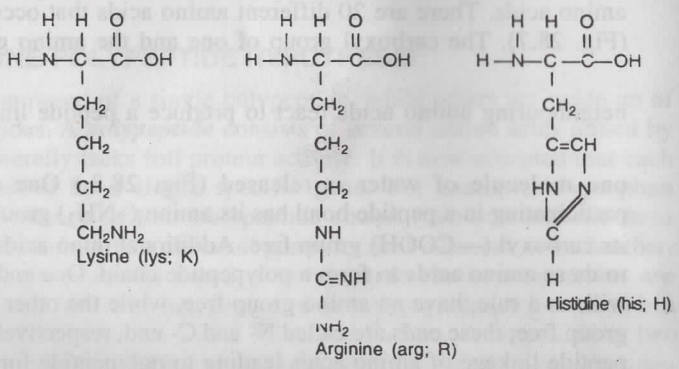
(2) Containing two carboxyl groups



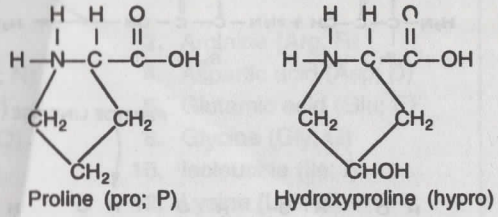
(3) Containing one amino and one amide group



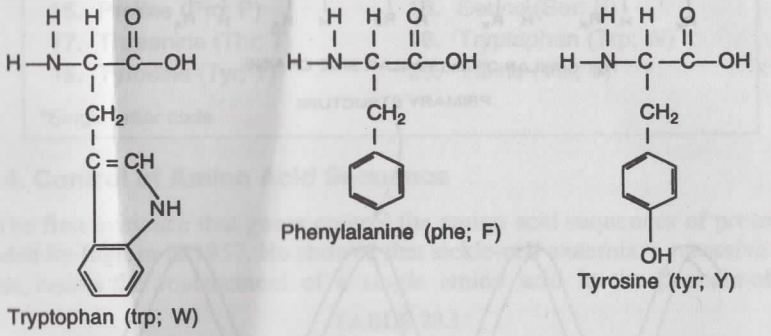
(4) Basic amino acids



(5) Cyclic amino acids



(6) Aromatic amino acids



(7) Sulphur containing amino acids

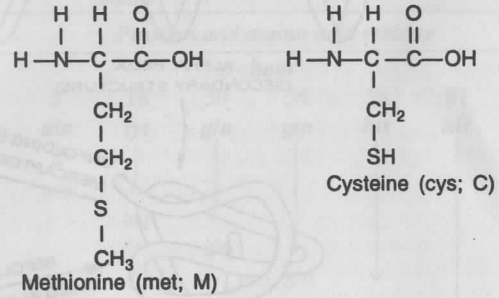


Fig. 28.7. Chemical structure of the common amino acids found in proteins. The letters within parenthesis are the common abbreviations representing the respective amino acids, while the single letters are the alphabetical codes for the amino acids that are used in writing amino acid sequences of proteins.

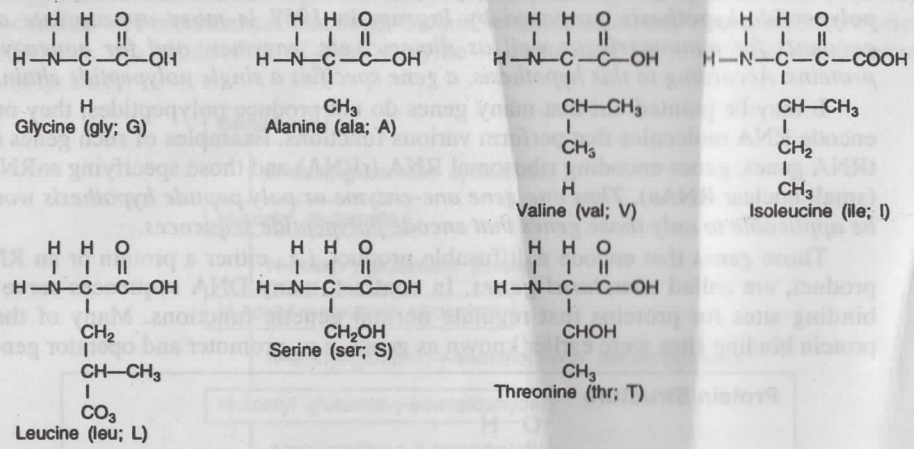
28.7.2. Primary Structure of Proteins

The structural and functional properties of a protein molecule depend on the proportion of the different amino acids present and, more particularly, on the precise sequence of the amino acids present in the molecule. Each protein has a specific amino acid sequence. A change of even a single amino acid may alter the function of a protein molecule. The sequence of amino acids in a protein molecule is known as its ***primary structure.***

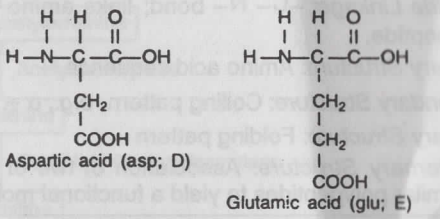
28.7.3. Secondary and Higher Structures

Most protein molecules are coiled into a right-handed helix, the ***alpha helix***, which gives them their ***secondary structure*** (Fig. 28.8). In most of the proteins, the

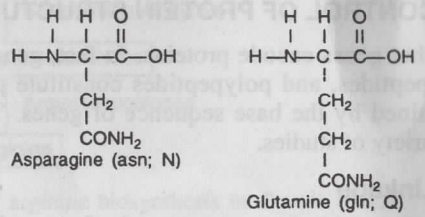
(1) Containing one amino (—NH_2) and one carboxyl (—COOH) group



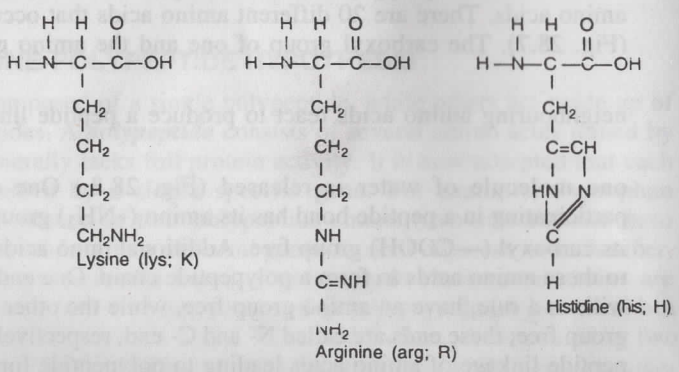
(2) Containing two carboxyl groups



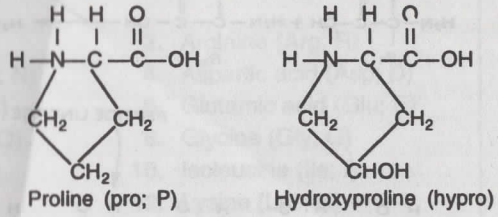
(3) Containing one amino and one amide group



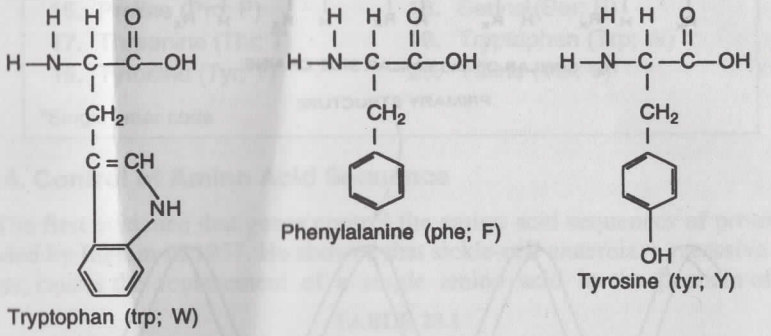
(4) Basic amino acids



(5) Cyclic amino acids



(6) Aromatic amino acids



(7) Sulphur containing amino acids

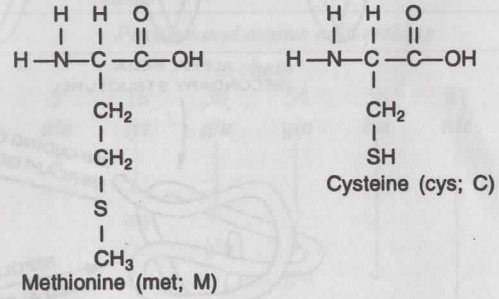


Fig. 28.7. Chemical structure of the common amino acids found in proteins. The letters within parenthesis are the common abbreviations representing the respective amino acids, while the single letters are the alphabetical codes for the amino acids that are used in writing amino acid sequences of proteins.

28.7.2. Primary Structure of Proteins

The structural and functional properties of a protein molecule depend on the proportion of the different amino acids present and, more particularly, on the precise sequence of the amino acids present in the molecule. Each protein has a specific amino acid sequence. A change of even a single amino acid may alter the function of a protein molecule. The sequence of amino acids in a protein molecule is known as its ***primary structure.***

28.7.3. Secondary and Higher Structures

Most protein molecules are coiled into a right-handed helix, the ***alpha helix***, which gives them their ***secondary structure*** (Fig. 28.8). In most of the proteins, the

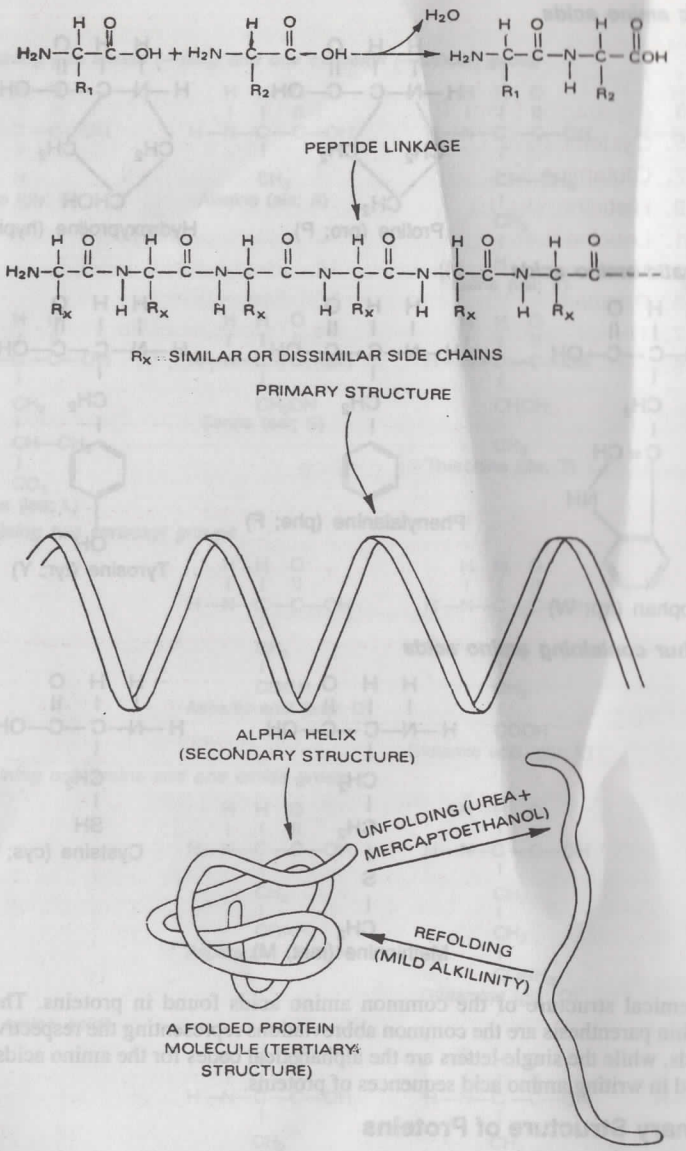


Fig. 28.8. Primary, secondary and tertiary structures of a protein molecule. In polypeptide chains a (C-C-N) *n* backbone is present; here *n* is equal to the number of amino acid residues in the chain.

alpha helix is folded at various places to produce a compact spheroid molecule giving rise to the ***tertiary structure*** of protein molecule. *The secondary and tertiary structures of a protein molecule are determined by its primary structure.* In addition, many proteins consist of two or more polypeptide chains, which gives them their ***quaternary structure.*** The different polypeptide chains are held together by noncovalent and, sometimes, covalent bonds.

Amino Acids

1. Alanine (Ala; A*)
2. Arginine (Arg; R)
3. Asparagine (Asn; N)
4. Aspartic acid (Asp; D)
5. Cysteine (Cys; C)
6. Glutamic acid (Glu; E)
7. Glutamine (Gln; Q)
8. Glycine (Gly; G)
9. Histidine (His, H)
10. Isoleucine (Ile; I)
11. Leucine (Len; L)
12. Lysine (Lys; K)
13. Methionine (Met; M)
14. Phenylalanine (Phe; F)
15. Proline (Pro; P)
16. Serine (Ser; S)
17. Threonine (Thr; T)
18. Tryptophan (Trp; W)
19. Tyrosine (Tyr; Y)
20. Valine (Val; V)

*Single letter code

28.7.4. Control of Amino Acid Sequence

The first evidence that genes control the amino acid sequences of proteins was provided by Ingram in 1957. He showed that sickle-cell anaemia, a recessive genetic disease, causes the replacement of a single amino acid in the β chain of human

TABLE 28.1

Amino acid changes and their positions in some of the haemoglobin variants for A and B chains

Haemoglobin	*Position and amino acid residue*							
	α chain							
	1	5	16	30	54	58	87	116
Normal	*val*	*ala*	*lys*	*glu*	*gin*	*his*	*his*	*glu*
Hbj Toronto		asp						
Hbl			asp					
HbG Chinese				gln				
Hb Mexico					glu			
HbM Boston						tyr		
Hbm Iwate							tyr	
Hbo Indonesia								lys
	β chain							
	1	6	16	26	46	63	67	87	143
Normal	*val*	*glu*	*gly*	*glu*	*gly*	*his*	*val*	*thr*	*his*
HbC	lys								
HbS		lys	val						
Hb Baltimore			asp						
HbE				lys					
HbK Ibadan					gln				
HbM Saskatoon						trp			
Hb Sydney							ala		
HbD Ibadan								lys	
HbK Kenwood									asp

haemoglobin. RBCs (red blood cells) of sickel cell anaemia patients assume a typical sickle shape under conditions of low oxygen concentration. This is because the haemoglobin in these cells forms crystals under low oxygen concentration.

Normal adult haemoglobin (HbA) has net negative charge, while sickle cell haemoglobin (HbS) has a net positive charge. Ingram showed that the two heamoglobins, *viz.* HbA and HbS, differ for a single amino acid at position 6. (In a polypeptide chain, position 1 is conventionally the N-terminal amino acid of the chain and is written at the extreme left of the sequence.) He showed that the glutamic acid present at position 6 in HbA is replaced by valine in HbS. *This clearly showed that a gene mutation acts by changing the amino acid sequence of the polypeptide encoded by the gene.* A large number (over 100) of variants of human haemoglobin are known; some of these are summarised in Table 28.1. In each case, a single amino acid of either α or β chain has been replaced.

28.8. COLINEARITY OF POLYNUCLEOTIDES AND POLYPEPTIDES

In 1958, Crick proposed that DNA determines the sequence of amino acids in polypeptides. Thus a given sequence of nucleotides should specify a given amino acid sequence. Therefore, change in the nucleotide sequence should produce a corresponding change in the amino acid sequence of the concerned polypeptide. Further, if several changes occur separately in a single polynucleotide sequence, the order of these changes should match perfectly with the order of changes produced by

Genes and Amino Acid Sequences of Proteins

- Genes control amino acid sequences of proteins.
- Ingram (1957) showed sickle cell haemoglobin (HbS) to have Val at position 6 in place of the usual Glu (found in HbA).
- Crick (1958) proposed that DNA determines amino acid sequence of proteins.
- *Colinearity of Polynucleotides and Polypetides*: Linkage map of different mutations in a gene matches precisely the location of amino acid substitutions produced by them in the concerned protein.
- The base sequence of a gene predicts precisely the amino acid sequence of the protein encoded by it.

them in the amino acid sequence of the concerned polypeptide. These expectations summarise the concept of ***colinearity of polypeptide and polynucleotide sequences***.

For example, the α polypeptide of tryptophan synthetase of *E. coli* has 267 amino acids and is encoded by *tryptophan synthetase A* gene. Yanofsky and coworkers prepared a linkage map of different mutations in this gene. They compared this map with the locations of amino acide changes produced by these mutations in the α chain. There was a perfect match between the order in which the mutations were located in the gene and the location in α chain of the amino acid changes produced by them (Fig. 28.9). This confirmed the colinearity of polynucleotide and polypeptide sequences proposed by Crick.

The complete nucleotide sequences of RNA phage MS2 and the amino acid sequences of the three proteins produced by MS2 have been determined. The amino acid sequence in the protein shows a complete correspondance with the nucleotide

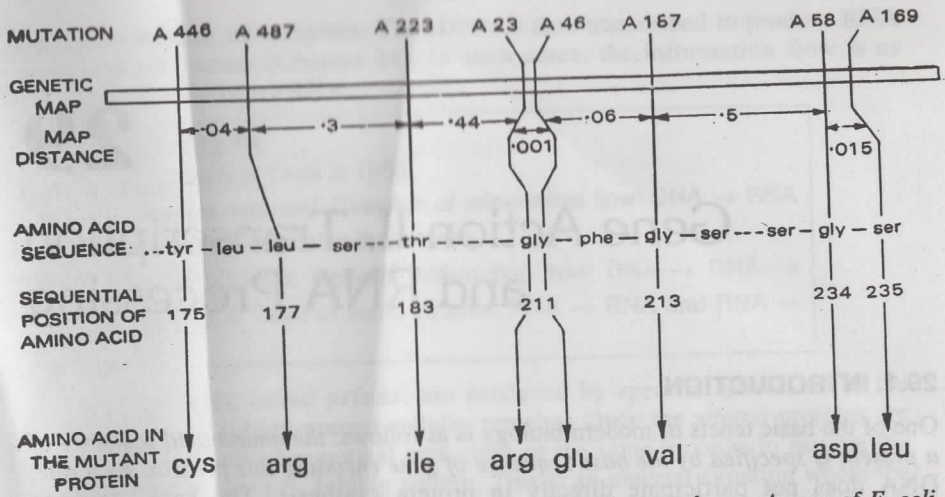


Fig. 28.9. Colinearity of mutations in a part of the *tryptophan synthetase A* gene of *E. coli* and the resulting amino acid changes in the α polypeptide of tryptophan synthetase.

sequence of the RNA genome as expected on the basis of genetic code (Chapter 30). This provides a conencing evidence in favour of the colinearity hypothesis.

QUESTIONS

1. 'Genes produce their effects by controlling the reactions of different biochemical pathways.' Comment giving appropriate examples.
2. Explain the one-gene-one-enzyme hypothesis and discuss its validity.
3. Briefly define the primary, secondary and tertiary structures of proteins. Briefly explain the relationship between amino acid sequence of a protein and the nucleotide sequence of the gene controling its production.
4. Write short notes on:
   (*i*) biochemical mutants of *Neurospora,*
   (*ii*) biosynthetic pathways,
   (*iii*) biochemical genetics,
   (*iv*) one-gene-one-polypeptide hypothesis,
   (*v*) one-gene-one enzyme hypothesis,
   (*vi*) peptide bond,
   (*vii*) protein structure.

29

Gene Action II. Transcription and RNA Processing

29.1. INTRODUCTION

One of the basic tenets of modern biology is as follows: *the amino acid sequence of a protein is specified by the base sequence of gene encoding this protein*. As a rule, DNA does not participate directly in protein synthesis. The base sequence information present in the gene (DNA) is copied into an RNA molecule, which directly participates in protein synthesis and provides information for amino acid sequence of the protein. This RNA molecule is called ***messenger RNA*** or ***mRNA***. The process of production of RNA copy of a DNA sequence is called ***transcription;*** this reaction is catalyzed by ***DNA-directed RNA polymerase***, or simply ***RNA polymerase***. The existence of mRNA was first shown by Spiegelman and coworkers in 1961, and RNA polymerase activity was first discovered by Wess in 1960.

29.2. THE CENTRAL DOGMA

In 1958, Crick proposed that the information present in DNA (in the form of base sequence) is transferred to DNA (*through replication*) and to RNA (*by transcription*), and from RNA to protein (*through translation*; Chapter 30). Thus the direction of information flow in biological systems is as follows DNA → RNA → protein; this is called the ***central dogma*** of biology (Fig. 29.1).

But it is now known that in many RNA viruses, genetic RNA replicates directly to produce RNA copies of itself; this reaction is catalyzed by RNA replicase (Chapter 24). In this case, information flows from RNA to RNA. In some other RNA viruses, the RNA genome is first reverse-transcribed into DNA (= cDNA); this reaction is

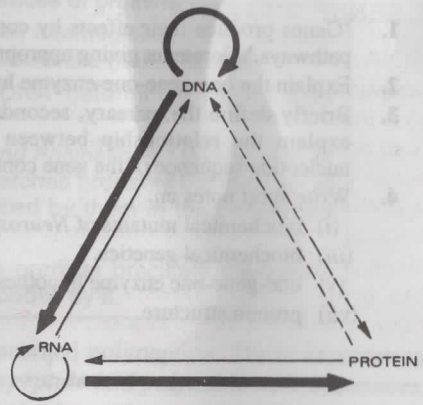


Fig. 29.1. The direction of flow of information among DNA, RNA and protein. Dark lines show the general direction of information flow, the light lines depict the situation observed in cells infected by RNA viruses, while the broken lines indicate the relationships not yet observed and most likely non-existent.

catalyzed by reverse transcriptase. The cDNA is then transcribed to produce RNA genomes of the viruses (Chapter 24). In such cases, the information flow is as follows : RNA → DNA → RNA.

Central Dogma

- Proposed by Crick in 1958.
- *Original proposal*: Direction of information flow: DNA → RNA → Protein. But not in the reverse.
- *Current Status*: General information flow: DNA → RNA → protein. In Case of some viruses: RNA → RNA and RNA → DNA.

Certain proteins, called ***prions***, are produced by specific alterations in the folding patterns of certain normal cellular proteins. Once the altered proteins are produced, they induce other normal molecules of the same proteins to assume the 'new' or altered shape, *i.e.*, folding pattern. This is an unusual situation where the change in folding pattern is transmitted from a protein molecule to another protein molecule. But even in this case, there is no transmission of amino acid sequence information from one protein to the other.

Thus central dogma may be summarised as follows: sequence (base/amino acid) information is transmitted from nucleid acid to nucleic acid and from RNA to protein; it is *not transmitted* from protein to nuclic acid.

29.3. SENSE AND ANTISENSE STRANDS

All RNA, except the genomes of many RNA viruses, is produced through transcription. During transcription, only one strand of a DNA molecule (= double helix) is transcribed; this strand is called ***antisense strand*** or ***template strand*** and the RNA so produced is termed as ***sense RNA***. The other strand of the DNA duplex is known as ***coding strand*** or ***sense strand***. Therefore, the base sequence of sense RNA is complementary to that of the antisense strand of the DNA molecules, but is the same as that of the sense strand (except that in the please of *T* in the sense strand, *U* will be present in the sense RNA).

As in the case of DNA replication, RNA chain elongates in the 5′ → 3′ direction, and a free 3′ – OH group is essential for the addition of new ribotides to the growing RNA chain. Therefore, as a rule, the orientation of the antisense strand is P – 3′ → 5′, where *P* is the promoter sequence of the concerned gene (Section 4). As a result, transcription proceeds from promoter toward the 5′-end of the antisense strand. As a matter of fact, *it is the orientation of a given strand of a gene in relation to the promoter of the gene that determines if it will function as an antisense or a sense strand.*

29.4. TRANSCRIPTION UNIT

A ***transcription unit*** is that stretch or sequence of DNA that is transcribed into a single RNA molecule (Fig. 29.2). A typical transcription unit has (1) a promoter at its beginning (at the 3′-end of its antisense strand), (2) a startpoint, (3) a coding region and (4) a terminator sequence at its end. RNA polymerase binds to the promoter, and transcription begins at the startpoint; it progresses through the coding

sequence and ends around the terminator site. *In case of eukaryotes, each gene is a distinct transcription unit. But in case of prokaryotes, genes encoding the enzymes for a single biosynthetic pathway are usually clustered together into a single*

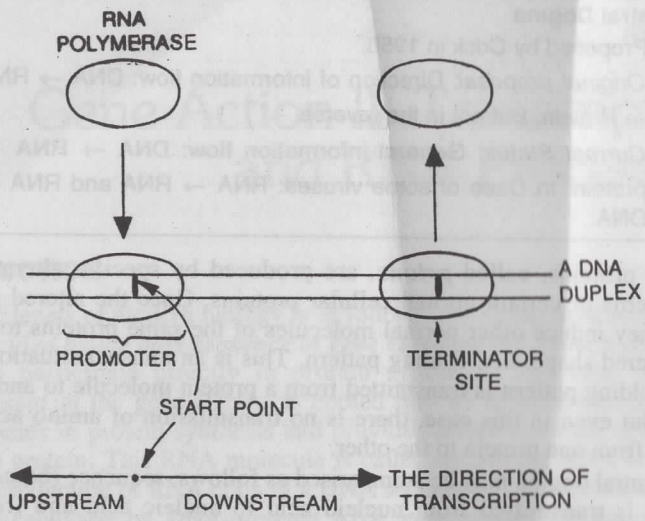


Fig. 29.2. A schematic representation of a transcription unit.

Transcription Unit

- The stretch of DNA transcribed into a single RNA molecule; an operon in prokarotes; a single gene in eukaryotes.
- Contains:
  1. a promoter,
  2. a startpoint,
  3. a coding region, and
  4. a terminator sequence.
- One strand of DNA duplex transcribed; called *antisense strand*; oriented *P*-3′ → 5′ (*P* = promoter).
- The strand complementary to antisense strand is *sense* strand; oriented *P* – 5′ → 3′.

regulatory unit called **operon** (Chapter 31); *each operon functions as a single transcription unit.* Conventionally, a transcription units is depicted as shown in Fig. 29.2, and transcription proceeds from left to right. The sequences located on the left of startpoint are called ***upstream sequences***, and the base positions are denoted sequentially as –1, –2, –3, etc. But the sequences located to the right of startpoint are termed as ***downstream sequences***, and the bases are sequentially designated as +1, +2, + 3, etc.

The RNA molecules obtained directly from transcription are called ***primary transcripts***; they contain all the sequences, beginning from the startpoint to the terminator site, present in the antisense strand. As a rule, primary transcripts are

unstable; they are either modified, *e.g.*, eukaryotic mRNAs, all rRNAs and tRNA, or are degraded (prokaryotic mRNAs) rapidly.

29.5. OPTIMAL PROKARYOTIC PROMOTER

Promoter is that sequence of a transcription unit where RNA polymerase binds and initiates transcription. In fact, a promoter sequence will support transcription of any DNA double helix that is joined to it correctly. Base sequences of a large number of prokaryotic promoters have been determined. *E. coli* promoters have the following 4 consensus sequences: (1) startpoint, (2) –10 sequence, (3) –35 sequence, and (4) the distance between –10 and –35 sequences (Fig. 29.3). A ***consensus sequence*** is a base sequence that is present in all the promoters (or other DNA sequences having the same function) but with some variation. In contrast, a ***conserved sequence*** is present, without any change, in all DNA sequence having the same function.

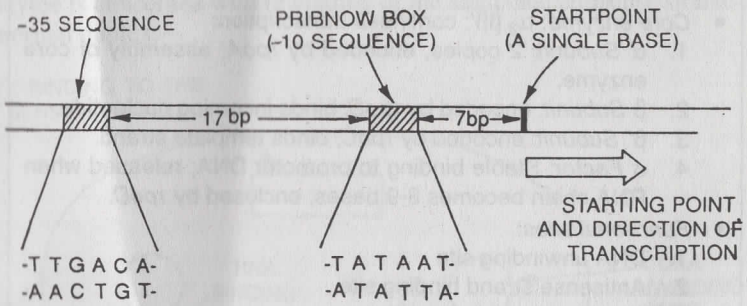


Fig. 29.3. A schematic representation of the optimal promoter, which has the following 3 consensus sequences : – 3 and – 10 sequences, and the startpoint.

29.5.1. Startpoint

Startpoint is the site within a promoter from which transcription begins. It is typically a single base within the promoter; in > 90% of the cases, startpoint is a purine in the antisense strand. Often startpoint is the A residue within the triplet CAT.

Optimal Prokaryotic Promoter

- RNA polymerase binds promoter and begins transcription.
- *E. coli* promoters have :
  1. A *startpoint*; in > 90% cases, a purine.
  2. A – 10 *sequence*; TATAAT; also called *Pribnow box.*
  3. A – 35 *sequence*; TTGACA; plays role in promoter recognition.
  4. *Distance between* – 10 and – *35 sequences*; usually 16-18 bp.

29.5.2. The –10 Sequence

The midpoint of –10 sequence is, on an average, 10 bp on the upstream (left) of the startpoint and has the consensus sequence TATAAT; this sequence is commonly known as ***Pribnow box***. This sequence is A·T–rich and, as a result, requires the minimum energy for strand separation or 'melting' during initiation of transcription.

29.5.3. The –35 Sequence

The midpoint of *–35 sequence* is, on an average, 35 bp upstream of the start-point; it has the consensus sequence TTGACA. This sequence plays an important role in promoter recognition by RNA polymerase.

29.5.4. The Distance between –10 and –35 Sequences

In 90% of promoters, the distance between –10 and –35 sequences is 16–18 bp; this distance seems to be critical for proper orientation of RNA polymerase during transcription initiation.

Bacterial RNA Polymerase

- *Holoenzyme*: α_2 β β′σ; binds promoter and initiates transcription; ~ 480,000 Da.
- *Core enzyme*: α_2 ββ′; continues transcription:
  1. α *Subunit*: 2 copies, encoded by *rpoA*; assembly of core enzyme.
  2. β *Subunit*: encoded by *rpoB;* binds incoming nucleotide.
  3. β′ *Subunit*: encoded by *rpoC*; binds template strand.
  4. σ *Factor*: Stable binding to promoter DNA; released when RNA chain becomes 8-9 bases; enclosed by *rpoD*.
- *Functional sites*:
  1. DNA unwinding site
  2. Antisense strand binding site,
  3. Sense strand binding site,
  4. DNA rewinding site.

29.6. BACTERIAL RNA POLYMERASE

In most bacteria, a single type of RNA polymerase is found. The complete RNA polymers molecule is called ***holoenzyme***; its has the following components: (1) the core enzyme and (2) the sigma (σ) factor. The holoenzyme may be symbolized as α_2 β β′σ.

29.6.1. Core Enzyme

The ***core enzyme*** can transcribe a DNA duplex after transcription has been initiated, but it can not initiate transcription at proper sites. The core enzyme has four polypeptides as follows: two α polypeptides, one β polypeptide and one β′ polypeptide. Therefore, core enzyme is symbolized as α_2 ββ′.

29.6.1.1. The α Subunit

Two copies of α polypeptide are found in each molecule of core enzyme. This polypeptide is encoded by gene *rpoA*. This subunit is required for assembly of the core enzyme; it may also play a role in promoter recognition.

29.6.1.2. The β Subunit

This subunit is present in one copy per core enzyme molecule, and is encoded by the gene *rpoB*. It is involved in binding with the incoming nucleotide to be added to the RNA chain during transcription.

29.6.1.3. The β′ Subunit

This subunit is also present is one copy per molecule of core enzyme, and is coded for by *rpoC*. It is involved in binding with the single–stranded template DNA.

The core enzyme carries out the following four functions, each function being performed presumably by a different site (Fig. 29.4).

1. ***DNA unwinding site*** unwinds the DNA duplex as it moves along the DNA being transcribed.
2. ***The site binding to antisense strand***, the strand being transcribed.
3. ***The site that binds to the sense strand***; this allows the antisense strand to remain single-stranded.
4. After a segment has been transcribed, the ***DNA rewinding site*** of core enzyme is concerned with rewinding of the sense and antisense strands into a normal duplex.

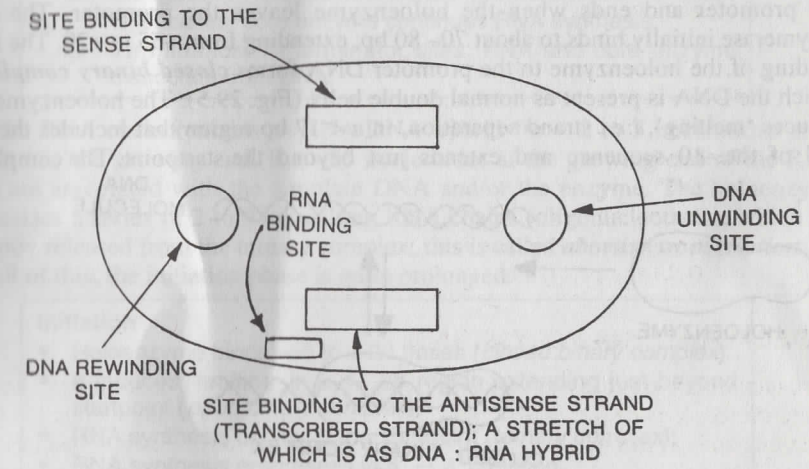


Fig. 29.4. A schematic representation of the different functional sites of the RNA polymerase core enzyme, the site binding to the template strand also contains the catalytic site.

29.6.2. Sigma Factor

The ***sigma factor*** is involved in stable binding of RNA polymerase holoenzyme specifically to promoter DNA. The σ factor is released from the core enzyme when the RNA chain reaches 8–9 bases following transcription initiation; the core enzyme now continues transcription. The *E. coli* σ factor (σ^{70}) is composed of a single polypeptide and is encoded by the gene *rpoD*. The chief function of σ factor is to ensure that the holoenzyme binds stably at only promoter sequences. In contrast, the core enzyme binds to any DNA sequence. Thus the ability of holoenzyme to recognize promoter sequences from others is due to the σ factor. The sigma factor does not bind DNA on its own; however, it does contact DNA when it is associated with the core enzyme.

Sigma factor has a decisive influence on the type of genes transcribed by the holoenzyme. The σ factor associates with the core enzyme during promoter recognition and transcription initiation, after which it separates from the core enzyme. An *E. coli* cell has ~7,000 core enzyme molecules all of which exist in association with DNA. The manner by which the holoenzyme molecules search out promoters is not known.

29.7. TRANSCRIPTION PROCESS

Transcription begins with the attachment of RNA polymerase holoenzyme to the promoter of a transcription unit and it ends when the core enzyme reaches the terminator site and dissociates from the DNA. The transcription process is divided into the following steps: (1) initiation, (2) elongation and (3) termination.

29.7.1. Transcription Initiation

The ***initiation*** phase begins with the binding of RNA polymerase holoenzyme to the promoter and ends when the holoenzyme leaves the promoter. The RNA polymerase initially binds to about 70– 80 bp, extending from –55 to +20. The initial binding of the holoenzyme to the promoter DNA forms ***closed binary complex***, in which the DNA is present as normal double helix (Fig. 29.5). The holoenzyme now induces 'melting', *i.e.*, strand separation, in a < 17 bp region that includes the right end of the –10 sequence and extends just beyond the startpoint. The complex of

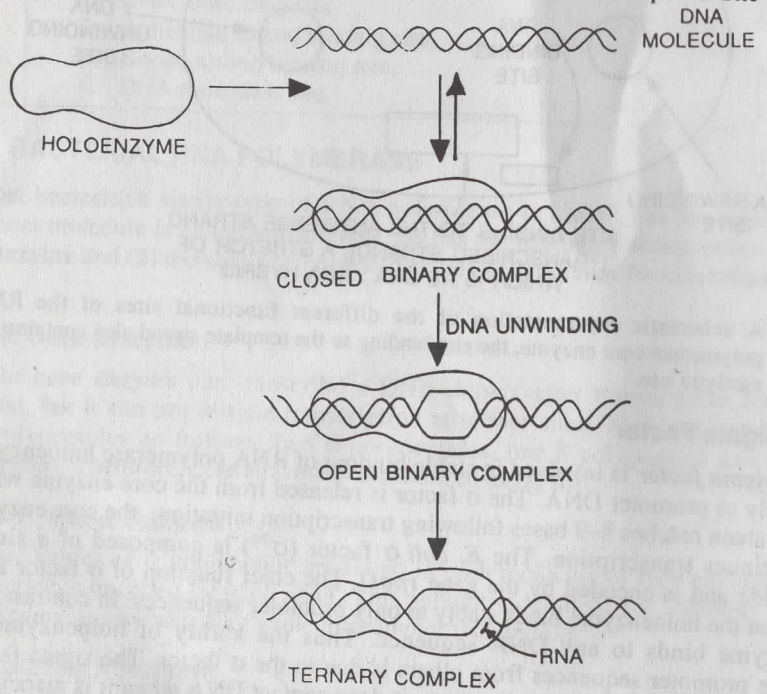


Fig. 29.5. Binding of RNA polymerase holoenzyme to DNA produces 'closed' and 'open' binary complexes, and the ternary complex.

melted DNA with the holoenzyme is known as ***open binary complex***. Usually, the formation of open binary complex is irreversible.

The holoenzyme now begins to transcribe the antisense strand; RNA synthesis begins at the startpoint. The nucleotides used are riboside 5′-triphosphates and RNA synthesis progresses in 5′ → 3′ direction. The enzyme permits the correct ribotides to align opposite the deoxyribotides of antisense strand. When the first two ribotides have been aligned, the enzyme catalyzes the formation of phosphodiester bond between them. The diribotide so formed remains associated with the template DNA strand and the enzyme; this gives rise to the ***ternary complex***. The holoenzyme sequentially adds the subsequent ribotides to the growing RNA chain.

Transcription Process

- Consists of 3 distinct phases:
  1. *Initiation*: Holoenzyme binds promoter and begins RNA synthesis.
  2. *Elongation*: Holoenzyme leaves promoter, and σ factor dissociates; core enzyme continues RNA synthesis.
  3. *Termination*: At terminator sites; RNA and core enzyme dissociate from DNA.

The RNA chain remains as RNA·DNA hybrid for ~ 2 to 3 nucleotides at its 3′-end, *i.e.*, the growing end. But ~25 nucleotides at the growing end of the RNA chain are associated with the template DNA and/or the enzyme. The holoenzyme synthesizes a series of 2 to 9 bases long RNA chains (oligonucleotides), which are promptly released from the ternary complex; this is called ***abortive transcription***. As a result of this, the initiation phase is quite prolonged.

Initiation

- Holoenzyme binds –55 to + 20 bases (*closed binary complex*).
- It induces 'melting' in < 17 bp region extending just beyond startpoint (*open binary complex*).
- RNA synthesis begins at the startpoint (*ternary complex*);
- RNA synthesis progresses in 5′ → 3′ direction.
- RNA chain is present as:
  1. *RNA•DNA hybrid*: terminal 2 – 3 ribotides
  2. *bound to DNA/RNA polymerase*: terminal 25 ribotides.
- *Abortive transcription*: A series of 2–9 bases long RNA chains synthesized and ejected before the holoenzyme leaves promoter.

29.7.2. Elongation

The ***elongation phase*** begins when the RNA polymerase leaves the promoter region and continues transcription of the template strand. Very early in this phase, the holoenzyme loses its σ factor, which may allow the core enzyme to leave the promoter. As the core enzyme moves along the DNA duplex, the regions downstream of the startpoint become progressively single–stranded. The core enzyme sequentially adds the correct ribotides to the 3′–end of RNA chain and, thereby, copies the base sequence of sense strand into an RNA copy. As the RNA chain elongates, its 5′-region progressively separates from the template strand. The

sense and antisense strands of the DNA become progressively free as the transcription proceeds; these single strands progressively reassociate to form double helix.

Elongation

- Holoenzyme leaves promoter.
- σ factor leaves holoenzyme around the same time.
- Core enzyme continues RNA synthesis till it reaches terminator site.
- *Transcription Bubble*: The transient single-stranded region of DNA; site of transcription.
- *Transcription Rate*: ~ 40 ribotides per second at 37° C.

Thus elongation involves progressive disruption of the double helix to generate transient single–stranded regions; the separated strands form a ***transcription bubble*** within which the template strand is transcribed. The transcription bubble moves progressively downstream of the startpoint and disappears when transcription terminates. Transcription bubble, as a rule is induced by and remains associated with RNA polymerase. The core enzyme moves along the DNA at an uneven pace as follows: the right (downstream) boundary of the RNA polymerase; does not move until several nucleotides have been added to the RNA chain, it then moves down 7-8 bp in one go. But the left boundary of the enzyme moves to the right regularly for every nucleotide added to the RNA. The overall rate of elongation for *E. coli* RNA polymerase is ~40 nucleotides per second at 37°C.

Termination

- *Rho-Independent Terminators*: RNA transcript forms a hairpin loop followed by a run of U residues; poly-U signals termination.
- *Rho-Dependent Terminators*: Require *rho* factor for termination; *rho* factor binds RNA transcript at a specific site and chases core enzyme; causes termination, when it catches the latter at termination site.

29.7.3. Termination

When the core enzyme reaches terminator site, (1) no further ribotides are added to the RNA chain, (2) the RNA chain dissociates from the template strand of DNA, and (3) the separated DNA strands reassociate to form a double helix. As a result, the transcription bubble disappears, and (4) the core enzyme dissociates from DNA. These events constitute the ***termination phase*** of transcription. The termination sites in prokaryotes have been classified into the following two groups: (1) *rho* – independent terminators, and (2) *rho*–dependent terminators.

29.7.3.1. Rho-Independent Terminator

Termination of transcription at these termination sites does not require *rho* factor; therefore, they are called ***intrinsic terminators***. The RNA transcript from such terminators forms a typical hairpin loop, which is followed by a run of ~6 U residues. The poly-U region probably signals the core enzyme to leave the DNA duplex. The actual termination may take place at any one of several positions toward the end of poly-U region.

29.7.3.2. Rho-Dependent Terminators

At these terminators, a polypeptide called ***rho-factor*** is required for transcription termination. Most likely, rho-factor binds the RNA transcript at a recognition site (~ 50 -90 bases long) that is located upstream of the termination site. Rho-factor moves along the RNA transcript at a faster rate than does RNA polymerase; if it catches up with the core enzyme at the termination site, it interacts with the enzyme to cause termination. Rho-factor functions as a hexamer. *E. coli* genome has few rho-dependent terminators.

29.8. TRANSCRIPTION IN EUKARYOTES

Transcription in eukaryotes is far more complex than that in prokaryotes, and the two differ with respect to many features (Table 29.1). A very brief and elementary description of transcription in eukaryotes is given below.

TABLE 29.1
A comparison of the transcription process in prokaryotes and eukaryotes

Characteristic	*Prokaryotes*	*Eukaryotes*
RNA Polymerase		
(*a*) Types	One in each species	Three; RNA polymerase I, II and III
(*b*) Molecular weight	~ 480,000 daltons	500,000 daltons
(*c*) Subunits	$\sigma_2\beta\beta'\sigma$	2 large subunits + <10 smaller subunits.
Transcription factors	Not known	Many different types of transcription factors; involved both in transcription and its regulation
Promoter	A simpler and relatively smaller sequence	A relatively larger sequence; has variable modules
Enhancer	A part of the promoter; may act in a manner similar to enhancer	Present at variable distances from the promoter
Transcription initiation	The holoenzyme binds to promoter and initiates transcription	Transcription factors first bind to promoter, then RNA polymerase associates with them and initiates
Transcription complex		
(*a*) Composition	Core RNA polymerase + σ factor	RNA polymerase + transcription factors
(*b*) Separation of components	σ factor dissociates from the core enzyme after initiation	Transcription factors dissociate when transcription is initiated
End product	Polycistromic RNA transcripts	Monocistronic RNA transcripts
Regulation		
(*a*) Type	Mainly negative control: inducible or repressible; positive control is less common	Primarily positive, rarely negative; inducible control
(*b*) Genes under the same regulation	Located in a single operon forming a single transcription unit; regulation results since all of them are controlled by a single promoter	Located at divergent sites; regulation results since all of them have the same response elements

Eukaryotic RNA Polymerases

- Have 2 **large and** < 10 small subunits; 5,00,000 Da.
- Three **types:**
  1. *RNA polymerase I*: 50 - 70% of total activity; rRNA genes.
  2. *RNA polymerase II*: 20 – 40% of activity; mRNA encoding genes.
  3. *RNA polymerase III*: ~10% of activity; tRNA and 5S RNA genes.

29.8.1. RNA Polymerase

All eukaryotes possess three types of RNA polymerases called RNA polymerase I, II and III (Table 29.2). ***RNA polymerase I*** is located in nucleolus, and *is responsible for transcription of genes for rRNA*; it is responsible for 50-70% of the activity in eukaryotes. ***RNA polymerase II*** is located in the nucleoplasm, constitutes 20-40% of total activity and *transcribes all the genes that produce mRNA*. ***RNA polymerase III*** also occurs in nucleoplasm, provides ~10% of total polymerase activity, and *transcribes tRNA and 5S RNA genes*. Each of these polymerases is a large protein having 2 large and 10-12 small subunits.

TABLE 29.2
Characteristics of the three RNA polymerases of eukaryotes*

Characteristic	*RNA polymerase I*	*RNA polymerase II*	*RNA polymerase III*
Subunits	2 + <10	2 + <10	2 + <10
(*a*) Large	2	2	2
	The largest subunit in each case is homologous to the β′ subunit of *E. coli*		
	The second subunit homologous to *E. coli* β subunit		
(*b*) Small	<10	<10	<10
	Three subunits related to that of RNA polymerase II	One subunit has homology with the *E. coli* α subunit	Three subunits related to that of RNA polymerase II
Target genes	Ribosomal RNA genes	mRNA genes	tRNA and 5S rRNA genes
Location	Nucleolus	Nucleoplasm	Nucleoplasm
Sensitivity to α-amanitin	Not sensitive	Very sensitive	Species-specific; much less sensitive
Promoters	Upstream of the startpoint	Upstream of the startpoint	Downstream of the startpoint
Level of activity	Most prominent (5-70%)	Second to polymerase I (20-40%)	Minor activity (~10%)
Carboxy-terminal domain (CTD)	Absent	Present (in the largest subunit)	Absent
Acillary factors	Two (UBF1, SL1)	Many (basal, upstream, regulatory)	Three

*In addition, mitochondria and plastids have their own RNA polymerases.

In addition, mitochondria and plastids present in eukaryotic cells contain RNA polymerases that are analogous to phage and bacterial polymerases.

29.8.2. Transcription Factors

Transcription factors are those proteins that are essential for transcription initiation, but they are not a part of RNA polymerases. A large number of transcription factors function with RNA polymerase II; they are divided into the following three groups: (1) basal, (2) upstream and (3) regulatory transcription factors.

Transcription Factors

- Proteins essential for transcription in eukaryotes.
- Three types:
  1. *Basal Factors*: Required for initiation; TFIIA, B, D, E, F, H, J.
  2. *Upstream factors*: Bind to upstream sequences; increase initiation frequency.
  3. *Regulatory factors*: Bind to response elements; regulate transcription.

29.8.2.1. Basal Transcription Factors

These factors are required for transcription initiation at all the promoters. They join RNA polymerase II to form a complex around the startpoint, and determine the site of transcription initiation. The different basal factors are as follows: (1) TFIIA, (2) TFIIB, (3) TFIID, (4) TFIIE, (5) TFIIF, (6) TFIIH, and (7) TFIIJ.

29.8.2.2. Upstream Transcription Factors

These transcription factors are found in all cell types and bind to specific short sequences located upstream of the startpoint. These factors act on any promoter having the appropriate sequence and increase the frequency of initiation. Protein SP1 is an example of upstream transcription factors.

29.8.2.3. Regulatory Transcription Factors

These factors function just like upstream factors, but they have a regulatory role. They are produced or activated at specific times or in specific tissues. As a result, they control transcription of the concerned genes (Chapter 31).

29.8.3. Promoters

Eukaryotic promoters are defined as regions that can support transcription at normal efficiency and with the proper control. The organization of promoters for the three types of RNA polymerases differ markedly. The promoters for RNA polymerase II usually have the following modules or functional sequences (Fig. 29.6).

29.8.3.1. Initiator (Inr)

It is the region that contains startpoint and has the general form Py_2CAPy_5. It is recognized by RNA polymerase II.

29.8.3.2. TATA Box

It is a consensus sequence of 8 bases (TATAAAAA), having only A·T base

pairs, and is the only consensus sequence that occupies a fixed position in the promoter at –25. It is usually surrounded by a G·C-rich sequence. A minority of promoters do not have a TATA box.

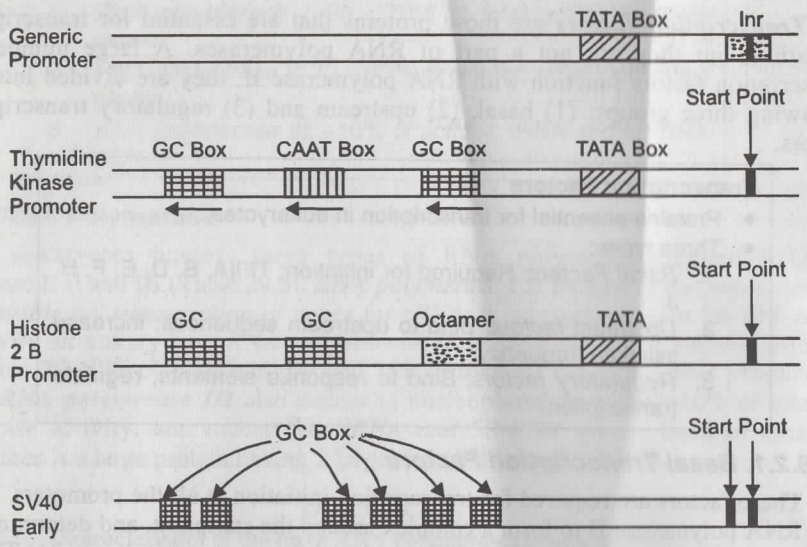


Fig. 29.6. Arrangement of the different modules in some selected eukaryotic promoters. The TATA box controls the choice of startpoint. The direction of arrows below the modules in the thymidine kinase promoter indicates their orientatin in the promoter.

29.8.3.3. CAAT Box

This sequence is located ~80 bp upstream of the startpoint, has the consensus sequence GGCCAATCT, and it increases promoter strength.

Eukaryotic Promoters

- Sequences that support initiation at normal efficiency and with proper control.
- Promoters for RNA polymers II have:
  1. *Initiator* (*Inr*); Py_2CAPy_5; has startpoint.
  2. *TATA Box:* Sequence TATAAAAA; at – 25; *Inr* TATA box form generic promoter.
  3. *CAAT Box*: GGCCAATCT; usually ~80 bp upstream.
  4. *GC Box*: GGGCGG; at –90.
  5. *Octamer Sequence*: ATTTGCAT.
- Sequences 3, 4, 5 are upstream sequences.

29.8.3.4. GC Box

It has the consensus sequence GGGCGG, is usually located at –90 and may be present in several copies in a promoter.

29.8.3.5. Octamer Sequence

It has the consensus sequence ATTTGCAT and is recognized by more than one transcription factor, *e.g.*, Oct-1 and Oct-2.

Sequences *Inr* and TATA box constitute the ***generic promoter*** to which the basal transcription factors bind to form the ***initiation complex***. The other consensus sequences are upstream sequences to which upstream transcription factors bind; these factors most likely act directly on the basal transcription factors and increase the frequency of their assembly into an initiation complex.

None of the consensus sequences is uniquely essential for promoter activity, and the location of upstream sequences is variable. *The choice of startpoint seems to depend on the location of TATA box.*

Enhancers

- Increase function of promoters lying in their vicinity.
- Function in either orientation.
- Located on either side of promoter upto several kb away.
- Organization similar to that of promoters.
- Most likely function by increasing the concentration of transcription factors in the vicinity of promoters.

29.8.4. Enhancer

Some sequences or elements affect the activity of a promoter located in either direction and located upto several kilobases away from them; they are called ***enhancers***. An enhancer (1) does not function as a promoter, (2) functions in either orientation, and (3) is located at a variable positions in relation to the affected promoter. An enhancer can stimulate any promoter that is placed in its vicinity. Enhancers have a similar general organization as promoters, and some of the elements present in enhancers are also found in promoters. Therefore, an enhances may be regarded as a part of the promoter that is located at a considerable distance from the startpoint. An enhancer affects the promoter that is located nearest to it on the same DNA molecule. The essential role of enhancer seems to be to increase the concentration of some of the transcription factors in the vicinity of the promoter; this they achieve by binding to these factors.

29.8.5. The Process of Transcription

In eukaryotes transcription unit generally contains a single gene. Transcription termination occurs beyond the end of coding region; the mechanism is known for RNA polymerases I and III but not for RNA polymerase II. Transcription by RNA polymerase II is briefly described below.

29.8.5.1. Transcription Complex

The complex of proteins formed at the promoter that ultimately leads to transcription initiation is called ***transcription complex***; this is the first step in transcription initiation. It is formed as follows.

1. The first step in the initiation of transcription is the binding of basal transcription factor TFIID (transcription factor D for RNA polymerase II) to the TATA box (Fig. 29.7).

2. TFIIA now binds to the promoter immediately upstream of TFIID.
3. TFIIB is now able to bind the promoter just downstream of TFIID.
4. At this point TFIIF–RNA polymerase II joins the transcription factors bound at the promoter. TFIIF binds tightly to the RNA polymerase and brings it to the promoter; it also helps the enzyme to bind the promoter possibly by interacting with TFIIB.
5. Now TFIIE joins the transcription complex developing at the promoter.
6. After TFIIE, factor TFIIH and TFIIJ join the promoter.

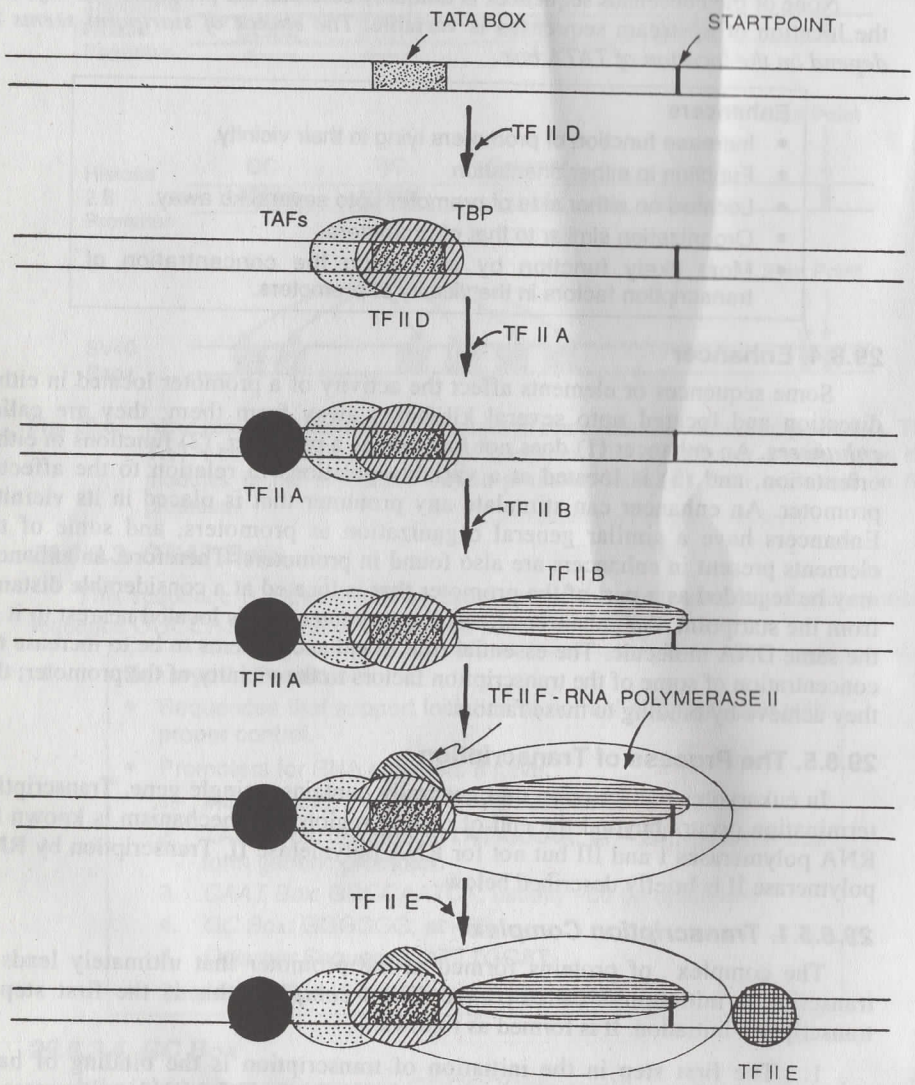


Fig. 29.7. The sequence of binding of transcription factors during the development of transcription complex at a generic promoter in eukaryuotes (RNA polymerase II).

Transcription Initiation
- Startpoint is usually an A.
- TFIID binds TATA box.
- TFIIA binds upstream of TFIID.
- TFIIB binds downstream of TFIID.
- TFIIF–RNA polymerase II binds promoter.
- TFIIE joins promoter.
- TFIIH and TFIIJ bind promoter.

The initiation complex is formed.
- DNA 'melting' induced around startpoint; TFIIE is lost.
- RNA polymerase begins transcription.
- Transcription factors released.

29.8.5.2. Transcription Initiation

Transcription begins at the startpoint, which is usually an A. A ***closed binary complex*** is formed when RNA polymerase II binds the promoter (item 4 above). The complex is then converted into an ***open binary complex*** in which the two strands of promoter DNA become separated locally; TFIIE is lost around this stage. A reaction requiring ATP hydrolysis may lead to the release of some more factors from the initiation complex. RNA polymerase now begins transcription at the startpoint, and presumably, the transcription factors are released around this stage.

The basal transcription factors are able to initiate transcription at a low level. Efficient initiation requires the effects of certain upstream and/or regulatory factors; these factors most likely interact with the basal factors and, thereby, enhance the frequency of formation of initiation complex.

Transcription Termination
- Terminator regions rather long with multiple terminator sites.
- In SV40 termination site has a hairpin loop followed by poly-(U).
- Mechanism of termination unknown.

29.8.5.3. Transcription Termination

It is likely that the transcription units have rather long terminator regions with multiple terminator sites. In the virus SV40, the termination site consists of a hairpin loop, followed by a stretch of U bases; this is similar to the intrinsic terminators of prokaryotes. In some transcription units, termination occurs >1,000 bp downstream of the site corresponding to the 3′-end of mature mRNAs. In such cases, the 3′-ends of mRNAs are generated by **cleavage following** transcription. The mechanism of termination is not well known.

Changes in Chromatin Structure
- DNA of active genes more exposed.
- Nucleosome organization disrupted temporarily in the region of transcription.

29.8.6. Changes in Chromatin Structure

An RNA polymerase engaged in transcription binds tightly to ~ 50 bp DNA; of

this ~12 bp are in unwound state. The DNA in transcriptionally active genes is more exposed than it is in the case of nontranscribed genes. It is likely that the nucleosomes are temporarily displaced as RNA polymerase passes by during transcription; they are reformed immediately after the enzyme goes past a nucleosome. It is possible that the histone octamer of nucleosomes may remain associated with the sense strand. (Alternatively only the $H3_2$. $H4_2$ tetramer may remain associated with the sense strand; H2A and H2B may join them later on.)

RNA Processing

- Production of functional RNA molecules from larger primary transcripts; usually by cleavage.
- *Ribosomal RNAs*: Primary transcript is methylated; sequential cleavage yields functional rRNA molecules.
- *tRNAs* : tRNA secondary structure helps in processing; tRNA produced by cleavage and trimming.
- mRNA : Processed in several ways:
  1. Self-splicing (group I introns).
  2. Self-splicing (group II introns).
  3. Splicing based on secondary structure.
  4. Splicing mediated by spliceosomes.

29.9. RNA PROCESSING

The functional rRNA and tRNA molecules in both prokaryotes and eukaryotes, and mRNAs in eukaryotes are produced from much longer primary transcripts. The process of modification, mainly through cleavage and/or splicing, of primary RNA transcripts so as to produce functional tRNA, rRNA and mRNA molecules from them is called ***RNA processing***. RNA processing is carried out by ***ribonucleases***, the enzymes that cleave RNA. They are of two types: (1) exoribonucleases and (2) endoribonucleases. ***Exoribonucleases*** remove one base at a time from one end of RNA molecule; all known enzymes trim RNA from the 3′-end. But ***endoribonucleases*** produce internal cuts in RNA molecules. Ribonucleases not only process primary transcripts, they also degrade functional RNA molecules; both exo- and endoribonucleases are involved in RNA degradation.

29.9.1. Processing of Ribosomal RNAs

In both eukaryotes and prokaryotes, functional rRNA molecules are produced by processing larger primary transcripts. In mammals, the primary transcript is a 45 S RNA molecule, which has one copy each of the 28 S, 18 S and 5.8 S RNAs (Fig. 29.8). Before processing, the primary transcript (1) undergoes methylation if ribose residues (~110 $-CH_3$ groups per RNA molecule), and it (2) binds to proteins to form a ribonucleoprotein complex. The transcript is now processed as follows. However, little is known about the enzymes involved.

1. First of all, the 5′ leader sequence is cleaved (Fig. 29.8).
2. A cleavage now occurs in the region between 18 S and 5.8 S sequences. The 3′-end of 18 S and the 5′-end of 5.8 S molecules are most likely generated by trimming.
3. A cleavage now separates the 5.8 S RNA from the 28 S RNA. The 5.8 S RNA base-pairs with a segment of the 28 S RNA.

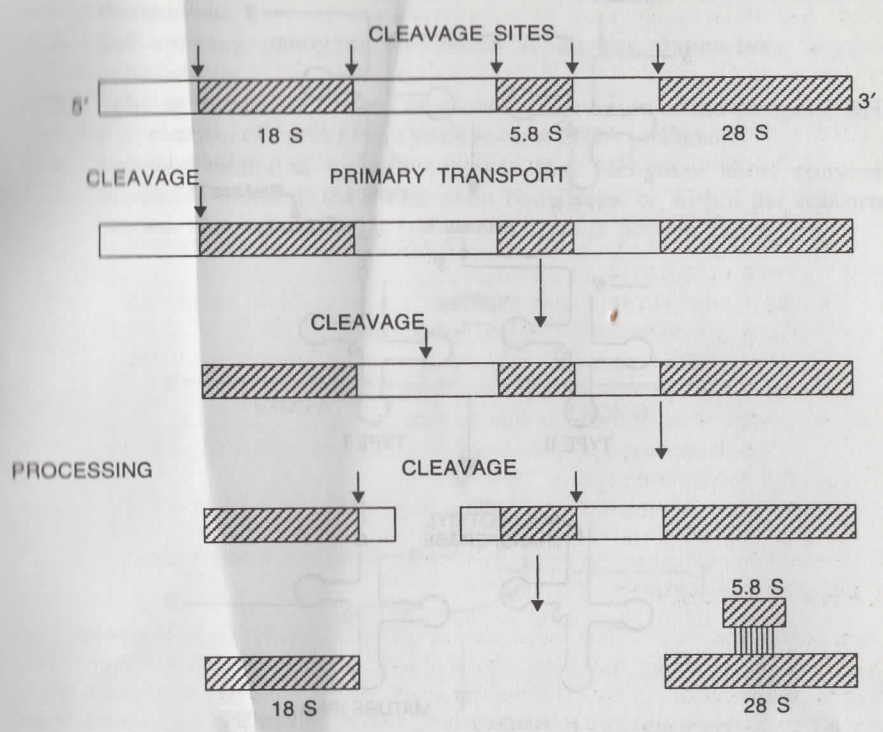


Fig. 29.8. Processing of the 45 S primary transcript of ribosomal RNA to release the 28 S, 18 S and 5.8 S rRNA molecules; the events described here are those observed in human HeLa cells (human cultured cells).

29.9.2. Processing of Transfer RNA

In both prokaryotes and eukaryotes, tRNA genes occur in clusters. For example, in *E. coli*, *tyrU* operon codes for 4 tRNAs and the protein EF-TU; the mRNA for EF-TU is generated by cleavage of the preceding tRNAs. The tRNA sequences in the precursor RNA (= the primary transcript) take up the same secondary structure as do the mature tRNA molecules; this seems to be important in their recognition by the ribonucleases during processing, which proceeds as follows (Fig. 29.9). The tRNA processing in eukaryotes is, however, not well known.

1. An endonuclease, ribonuclease *P* (RNAase *P*), cleaves the precursor RNA to produce the 5′–end of all the four tRNA molecules.
2. The 3′-termini of the tRNA molecules are generated by the exonuclease RNAase D. RNAase D sequentially degrades the 3′-end till it reaches the sequence CCA, the 3′-terminus in all tRNA molecules.

29.9.3. Processing of Messenger RNA

In eukaryotes, mRNAs of almost all the genes are produced by extensive processing. During mRNA processing, all the nitrons are excised from the primary

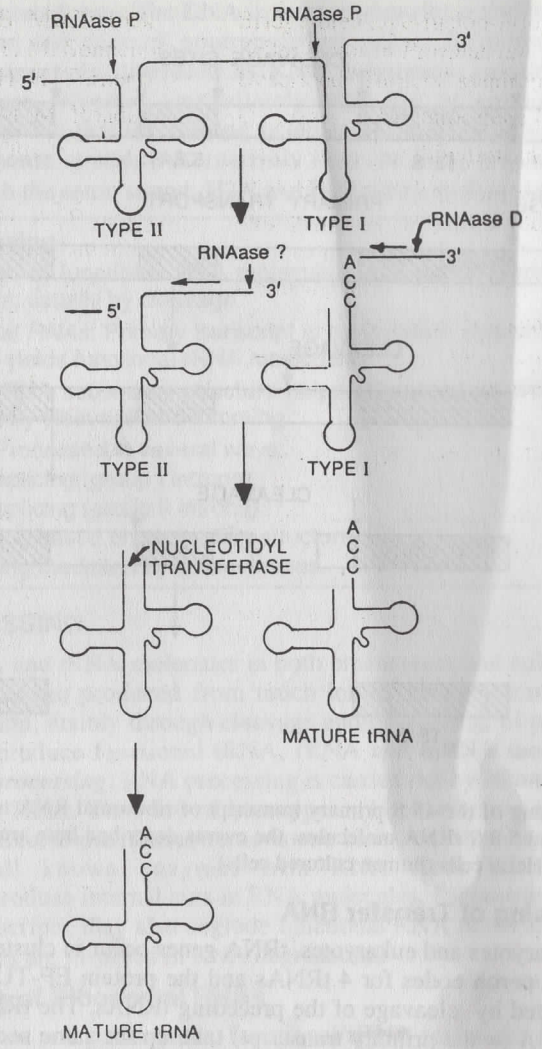


Fig. 29.9. Schematic representation of the processing of an RNA precursor containing one molecule of type I and one molecule of type II tRNA. The RNAase P is an endonuclease, while RNAase D and RNAase ? are exonucleases.

transcript and all the exons of a gene are joined together in the proper order; this process is called ***splicing***. The two boundaries between an intron and the two exons located at its two ends are called ***splicing junctions***. During splicing, the phosphodiester bonds at the two splicing junctions are cleaved and a new phosphodiester bond is formed between the two exons that begin at these junctions (Fig. 29.10). This excises the intron located between the two splicing junctions. Splicing can occur in the following four ways.

1. Self-splicing catalyzed by group I introns themselves, *e.g.*, in *Tetrahymena thermophila*.
2. Self-splicing catalyzed by group II introns themselves, *e.g.*, in mitochondria.
3. Splicing by enzymes based on secondary structure of the precursor RNA, *e.g.*, removal of introns from yeast nuclear tRNA precursors.
4. Splicing mediated by spliceosomes that recognize short consensus sequences either at the intron-exon boundaries or within the concerned introns. This is described in some more detail in Section 29.9.4.

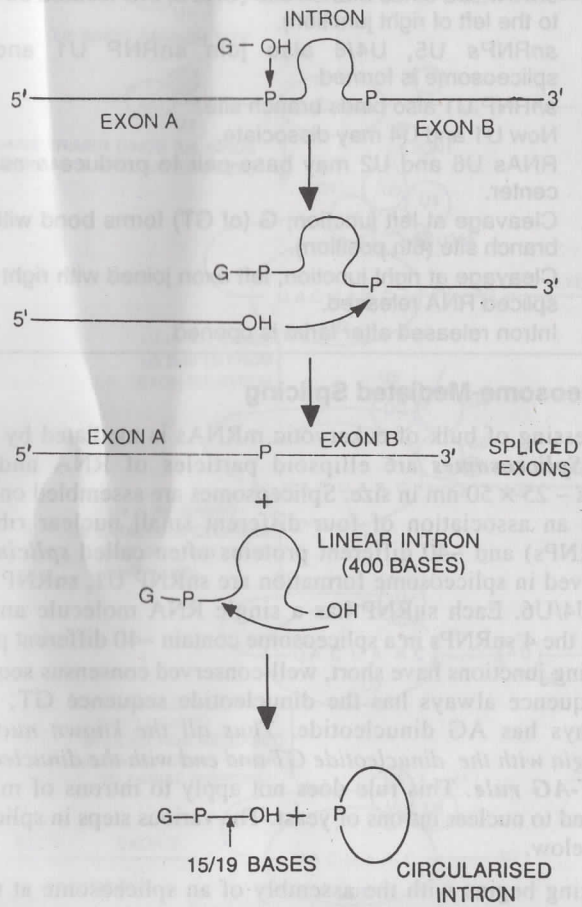


Fig. 29.10. Self-splicing catalysed by the intron of 35 S rRNA precursor of *Tetrahymena thermophila*. The intron is located in the 26 S rRNA sequence, and has the autocatalytic ability for splicing. The arrow indicates the point of transesterification.

Spliceosome - Mediated Splicing

- Processing of bulk of eukaryotic mRNA.
- *Spliceosome*: 25 × 50 nm ellipsoid particles formed by snRNPs U1, U2, U5 and U4/U6.
- *Left Splicing Junction*: Left end of intron has GT.
- *Right Splicing Junction*: Right end of intron has AG (thus, *GT–AG rule*).
- *Splicing Process*:
  1. *sn*RNP U1 binds left junction.
  2. *sn*RNP U2 binds branch site (UACUAAC located 30 bases to the left of right junction).
  3. *sn*RNP*s* U5, U4/6 also join snRNP U1 and U2; spliceosome is formed.
  4. *sn*RNP U1 also binds branch site.
  5. Now U1 and U4 may dissociate.
  6. RNAs U6 and U2 may base-pair to produce a catalytic center.
  7. Cleavage at left junction; G (of GT) forms bond with A of branch site (6th position).
  8. Cleavage at right junction; left exon joined with right exon; spliced RNA released.
  9. Intron released after lariat is opened

29.9.4. Spliceosome-Mediated Splicing

The processing of bulk of eukaryotic mRNAs is mediated by particles called spliceosome. ***Spliceosomes*** are ellipsoid particles of RNA and protein; each spliceosome is ~ 25 × 50 nm in size. Spliceosomes are assembled on precursor RNA molecules by an association of four different small nuclear ribonucleoprotein particles (snRNPs) and ~40 different proteins often called ***splicing factors***. The snRNPs involved in spliceosome formation are snRNP U1, snRNP U2, snRNP U5 and snRNP U4/U6. Each snRNP has a single RNA molecule and ~10 different proteins; thus the 4 snRNPs in a spliceosome contain ~40 different proteins.

The splicing junctions have short, well-conserved consensus sequences. The left consensus sequence always has the dinucleotide sequence GT, while the right junction always has AG dinucleotide. *Thus all the known nuclear introns of eukaryotes begin with the dinucleotide GT and end with the dinucleotide AG; this is known as* ***GT-AG rule***. This rule does not apply to introns of mitochondria and chloroplasts and to nuclear introns of yeast. The various steps in splicing reaction are summarised below.

1. Splicing begins with the assembly of an spliceosome at the left splicing junction. (1) snRNP U1 recognizes and binds to the dinucleotide GU (= GT in DNA) at the left junction. This enables (2) snRNP U2 to bind specifically to the UACUAAC consensus sequence (called ***branch site*** or ***TACTAAC box***) located within the intron about 30 bases to the left of the right junction of the intron. Now (3) snRNP U5 and snRNP U4/U6 join the snRNP U1 and U2; this completes spliceosome assembly (Fig. 29.11).

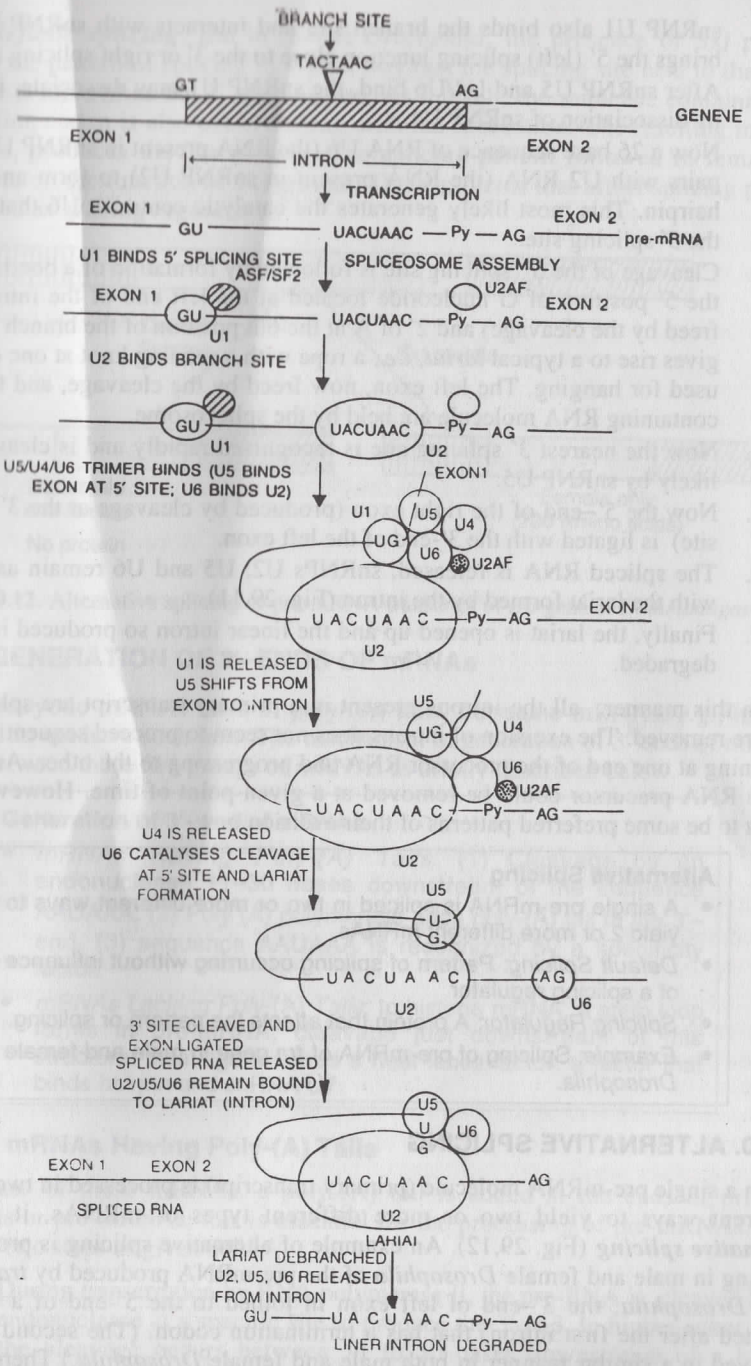


Fig. 29.11. Formation of spliceosome and splicing of pre-mRNA in eukaryotic nuclei. Py represents the pyrimidine of branch site.

2. snRNP U1 also binds the branch site and interacts with snRNP U2; this brings the 5′ (left) splicing junction close to the 3′ or right splicing site.
3. After snRNP U5 and U4/U6 bind, the snRNP U1 may dissociate, followed by dissociation of snRNP U4.
4. Now a 26 base sequence of RNA U6 (the RNA present in snRNP U6) base-pairs with U2 RNA (the RNA present in snRNP U2) to form an internal hairpin. This most likely generates the catalytic centre in U6 that cleaves the 5′ splicing site.
5. Cleavage of the 5′ splicing site is followed by formation of a bond between the 5′ position of G nucleotide located at the left end of the intron (now freed by the cleavage) and 2′ of A at the 6th position of the branch site; this gives rise to a typical ***lariat***, *i.e.*, a rope with a running knot at one end as is used for hanging. The left exon, now freed by the cleavage, and the lariat containing RNA molecule are held by the spliceosome.
6. Now the nearest 3′ splicing site is recognised rapidly and is cleaved most likely by snRNP U5.
7. Now the 5′–end of the right exon (produced by cleavage at the 3′ splicing site) is ligated with the 3′-end of the left exon.
8. The spliced RNA is released. snRNPs U2, U5 and U6 remain associated with the lariat formed by the intron (Fig. 29.11).
9. Finally, the lariat is opened up and the linear intron so produced is rapidly degraded.

In this manner, all the introns present in a primary transcript are spliced out, *i.e.*, are removed. The excision of introns does not seem to proceed sequentially, *i.e.*, beginning at one end of the precursor RNA and progressing to the other. Any intron of the RNA precursor could be removed at a given point of time. However, there seems to be some preferred patterns of their excision.

Alternative Splicing

- A single pre-mRNA is spliced in two or more different ways to yield 2 or more different mRNAs.
- *Default Splicing*: Pattern of splicing occurring without influence of a splicing regulator
- *Splicing Regulator*: A protein that affects the pattern or splicing.
- *Example*: Splicing of pre-mRNA of *tra* gene in male and female *Drosophila*.

29.10. ALTERNATIVE SPLICING

When a single pre-mRNA molecule (primary transcript) is processed in two or more different ways to yield two or more different types of mRNAs, it is called ***alternative splicing*** (Fig. 29.12). An example of alternative splicing is provided by splicing in male and female *Drosophila* of the pre-mRNA produced by *tra* gene. In male *Drosophila*, the 3′–end of left exon in joined to the 5′-end of a sequence (located after the first intron) that has a termination codon. (The second intron is removed in a similar manner in both male and female *Drosophila*.) Therefore, the resulting mRNA does not produce any protein. The splicing in males occurs without participation of a specific protein that affects the pattern of splicing; therefore, it is

called ***default splicing***. But in female *Drosophila*, the presence of Sxl protein changes the pattern of splicing in such a way that the splicing site next to that used in males is utilized as the 3′ splicing site. As a result, the sequence containing the termination codon is also excised along with the first intron. The resulting mRNA, therefore, produces the Tra protein. The splicing pattern followed in females is dependent on the function of Sxl protein. Those proteins that affect splicing pattern are termed as ***splicing regulators***.

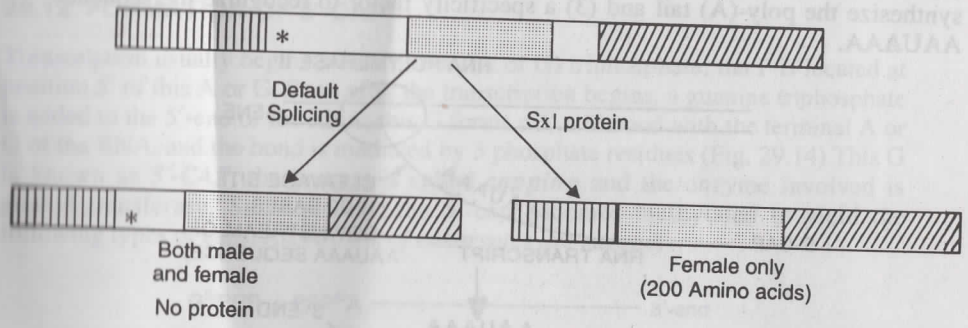


Fig. 29.12. Alternative splicing of pre-mRNA transcript of *tra* gene of *D. melanogaster*.

29.11. GENERATION OF 3′–ENDS OF mRNAs

Most eukaryotic mRNAs have 3′ poly-(A) tails, but some mRNAs, *e.g.*, histone mRNAs, do not have such tails. The mechanism of generation of 3′-ends of mRNAs differs between these two groups of mRNAs as briefly described below.

Generation of 3′- end of mRNAs

- *mRNAs Having Poly-(A) Tails*: (1) Cleavage by an endonuclease 11–30 bases downstream of the sequence AAUAAA. (2) Poly (A) polymerase adds poly-(A) tail to the 3′–end. (3) sequence AAUAAA is recognized by a specificity factor.
- *mRNAs Lacking Poly-(A) Tails*: In histone mRNA, a stem loop forms in pre-mRNA; cleavage just downstream of this structure; cleavage requires a heat labile factor, a factor that binds hairpin and snRNA U7.

29.11.1. mRNAs Having Poly-(A) Tails

In case of these mRNAs, a poly-(A) tail is added at the ends of primary transcripts or pre-mRNAs; this is essential for the processing of pre-mRNAs. This occurs in two steps as given below.

1. During transcription by RNA polymerase II, the pre-RNA is cleaved by an endonuclease at a specific site to produce the 3′-end. In higher eukaryotes, the cleavage occurs between 11 to 30 bases downstream of a highly conserved consensus sequence AAUAAA (Fig. 29.13); this sequence is needed for both cleavage and polyadenylation.

2. An enzyme poly-(A) polymerase now adds a poly-(A) tail to the newly generated 3′-end of the pre-mRNA. This enzyme uses ATP as substrate for the poly-(A) tail formation. First, a short poly-(A) tail of 10 residues is added; this is later extended to ~ 200 residues.

The cleavage and polyadenylation reactions require the following three proteins: (1) an endonuclease for cleavage of pre-mRNAs, (2) a poly(A) polymerase to synthesize the poly-(A) tail and (3) a specificity factor to recognize the sequence AAUAAA.

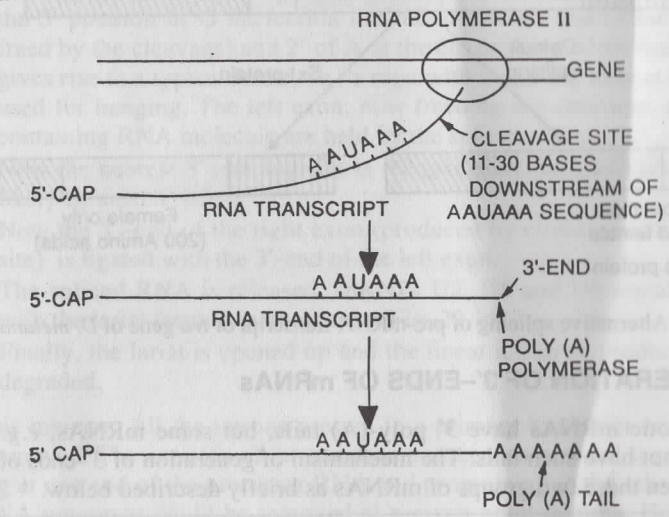


Fig. 29.13. The 3′-end of hnRNA is produced by a cleavage 11-30 bases downstream of a highly conserved consensus sequence AAUAAA. Poly(A) polymerase adds a poly(A) tail to this 3′-end.

29.11.2. mRNAs Lacking Poly(A) Tails

In such mRNAs, production of the 3′-ends depends on a secondary structure of the concerned pre–mRNA. In case of histone pre-mRNAs, this depends on the formation of a stem loop structure just upstream of the cleavage site. The cleavage requires the following 3 factors, in addition to the secondary structure in the pre-mRNA.

5′–Cap
- Forms soon after transcription begins.
- A GTP added at the 5′-end; a 5′-5′ bond is formed and is mediated by 3 P residues.
- *Cap O*: Position 7 of terminal G methylated; this is the only cap in lower eukaryotes.
- *Cap 1*: A -CH_3 group is added to the 2′ O of the second nucleotide of Cap O; predominates in higher eukaryotes.
- *Cap 2* : A –CH_3 is added to the 2′ O of the third base of Cap 1.
- All eukaryotic mRNAs are capped.

1. A heat-table factor with an unknown function.
2. A factor that binds to the hairpin just upstream of the cleavage site.
3. U7 snRNA, which base-pairs with a consensus sequence (CAAGAAAGA) present in the pre-mRNA just downstream of the cleavage site. The role of snRNA U7 in transcription termination and pre-mRNA cleavage is, however, not clear.

29.12. FORMATION OF 5′-CAP

Transcription usually begins with a purine (A or G) triphosphate; the P is located at position 5′ of this A or G. Soon after the transcription begins, a guanine triphosphate is added to the 5′-end of the RNA; this G forms a 5′ – 5′ bond with the terminal A or G of the RNA, and the bond is mediated by 3 phosphate residues (Fig. 29.14) This G is known as ***5′-CAP***, the process is called ***capping*** and the enzyme involved is guanyl transferase. Subsequently, the 5′-cap becomes methylated to yield the following types of caps. *All mRNAs of eukaryotes are capped.*

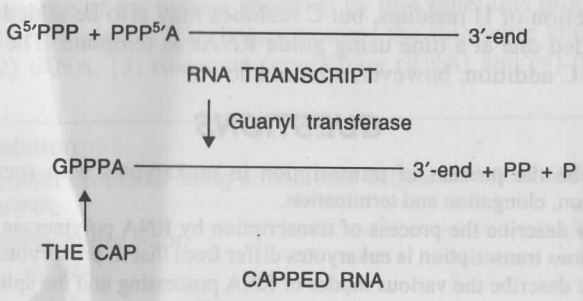


Fig. 29.14. Formation of 5′-cap in an RNA transcript.

1. ***Cap 0.*** In this type of cap, the 7 position of the terminal G that forms the cap is methylated. *This methylation occurs in all eukaryotes. In lower eukaryotes, this is the only cap that is found.*
2. ***Cap 1.*** This type of cap is produced from Cap 0 by addition of a methyl group to the 2′ O of the nucleotide next to the terminal G (the Cap). *Cap 1 is predominant in all higher eukaryotes.*
3. ***Cap 2.*** Another methyl group is added to the 2′ O of the third nucleotide from the terminal G (the Cap). This type of cap occurs in 10-15% of mRNAs.

RNA Editing
- Change in RNA base sequence after transcription.
- Extensive changes in trypanosome mitochondria
- Change directed by guide RNAs, encoded by separate genes.
- Usually, U residues added.
- U residues may be deleted.
- C residues may sometimes be added.

29.13. RNA Editing

It is widely accepted as an axiom that the base sequence of an mRNA represents only the sequence present in DNA from which it was transcribed. But in some cases, the base sequence of mRNA is changed after transcription; this is called ***RNA editing***. RNA editing may be confined to a single base or it may affect the entire mRNA. RNA editing is known in the case of mammalian mRNA for apolipoprotein B, where a single base is changed. More dramatic changes in mRNA base sequences occur in trypanosome mitochondria, and SV5 and measles viruses.

The site of editing is identified due to a relatively small sequence of ~50 bases in the vicinity of this site. The sequence information for editing is provided by ***guide RNAs***, which are encoded by genes different from those that encode the mRNAs to be edited. The guide RNA pairs with the pre-edited mRNA on both the sides of the site to be edited. The unpaired nucleotides of the guide RNA in this region provide information for, usually, addition of the complementary bases in the mRNA sequence.

Editing proceeds in the 3′→ 5′ direction of mRNA. RNA editing may involve addition or deletion of U residues, but C residues may also be added. U residues are most likely added one at a time using guide RNAs as template. The mechanisms of U deletion and C addition, however, are not known.

QUESTIONS

1. Describe the process of transcription in prokaryotes with special reference to initiation, elongation and termination.
2. Briefly describe the process of transcription by RNA polymerase II in eukaryotes. How does transcription is eukaryotes differ from that in prokaryotes.
3. Briefly describe the various modes of RNA processing and the splicing of mRNA.
4. Give a brief description of prokaryotic and eukaryotic promoters and list the various similarities and differences between them.
5. Briefly describe RNA processing in eukaryotes under the following heads:
   (*i*) spliceosome, (*ii*) generation of 3′-ends and (*iii*) 5′-capping.
6. Write short notes on the following:
   (*i*) central dogma, (*ii*) transcription unit,
   (*iii*) optimal prokaryotic promoter, (*iv*) promoter for RNA polymerase II,
   (*v*) bacterial RNA polymerase, (*vi*) sigma factor,
   (*vii*) transcription factors, (*viii*) enhancer,
   (*ix*) spliceosome, (*x*) alternative splicing,
   (*xi*) generation of 3′-ends of eukaryotic mRNAs, (*xii*) 5′-caps,
   (*xiii*) RNA editing.

30

Gene Action III. Translation

30.1. INTRODUCTION

Genes produce their phenotypic effects by specifying the amino acid sequence of specific proteins. As a first step in gene expression, the gene is transcribed to yield, ultimately, an mRNA. Then the base sequence of this mRNA is converted into the amino acid sequence of the protein specified by this gene; this process is known as ***translation***. The process of translation requires the following major components: (1) mRNA, (2) tRNA, (3) ribosome (containing rRNA) and (4) many translation factors.

Translation

- Protein synthesis using amino acid sequence information from mRNA.
- *Components*: Four.
  1. mRNA
  2. tRNA
  3. ribosomes
  4. Many translation factors.

30.2. MESSENGER RNA

Messenger RNA (mRNA) *is a single-stranded base-for-base complementary copy of the antisense stand of a prokaryotic transcription unit. But in case of eukaryotes, the pre-mRNA is processed by deleting introns, adding poly-(A) tails and by capping to yield mature mRNAs. The mRNA provides the information for amino acid sequence of the polypeptide specified by the concerned gene, or genes represented in the transcription unit. In prokaryotes translation proceeds along with transcription. But in eukaryotes, transcription and RNA processing are completed in the nucleus, then mRNA is transporated to cytoplasm where translation takes place.*

Generally, a single prokaryotic mRNA molecule codes for more than one polypeptide. Such an mRNA is known as ***polycistronic mRNA***. On the other hand, all eukaryotic mRNAs are ***monocistronic***, *i.e.*, code for protein specified by a single cistron. *All mRNA molecules have a translation initiation site close to their 5′-end and a chain termination site towards the 3′-end.* ***The chain initiation site*** consists of the codon AUG and an unknown secondary structure of mRNA. ***The chain termination*** site has one or two of the following three nonsense codons. UAA, UAG and UGA.

Messenger RNA

- Complementary RNA copy of antisense strand of gene.
- *Prokaryotic mRNA*: Polycistronic; short-lived (half-life ~2 min); translation simultaneous with transcription; main regions:
  1. *5′–leader*, untranslated, contains Shine-Dalgarno sequence (AGGAGG).
  2. Initiation codon (AUG).
  3. Coding region.
  4. Termination codon (UAA, UAG, UGA).
  5. Intercistronic region.
  6. 3′- trailer, not translated.

In prokaryotes, ribosomes attach to the 5′-end of mRNA as soon as transcription begins. A bunch of ribosomes moves along a single mRNA molecule adding ~15 amino acids/second to the polypeptide chain, almost the same speed at which RNA polymerase transcribes the mRNA. The group of ribosomes together with the single mRNA molecule they are translating is called ***polysome***. Bacterial mRNAs are ***short-lived***, the average functional half-life being ~2 min; their degradation closely follows translation.

In eukaryotes, mRNAs usually have functional half life of about 6 hours; for some mRNAs it may be upto 24 hr. The mRNA surviving more than 24 hr and upto 4 days are called ***moderately long-lived***. But mRNAs stored in seeds of plants survive for years (lotus seeds are known to germinate after 1,700 years); such mRNAs are called ***long-lived mRNAs.***

30.2.1. Prokaryotic mRNAs

In prokaryotes, all the genes present in one operon are transcribed into a single mRNA molecule. Therefore, all operons that have more than one gene yield polycistronic mRNAs. Both prokaryotic and eukaryotic mRNAs have the following regions: (1) ***a 5′ leader sequence*** that is not translated, (2) the ***coding region***, which begins with a translation ***initiation codon*** (ordinarily, AUG) and ends with a translation ***termination codon***, and (3) a nontranslated ***3′ trailer*** (Fig. 30.1). In case

Eukaryotic mRNA

- Produced by splicing.
- Has regions:
  1. 5′-cap, serves as ribosome binding site.
  2. 5-leader, untranslated.
  3. Initiation codon [AUG; context: GCCA (or G) CC AUG G].
  4. Coding region.
  5. Termination codon.
  6. Untranslated 3′–trailer of upto 1,000 bases.
  7. Poly-(A) tail of ~ 200 bases (in 70% of mRNAs).
- Short-lived (half–life < 24 hr.), moderately long-lived (> 24 hr) or long-lived (in seeds).

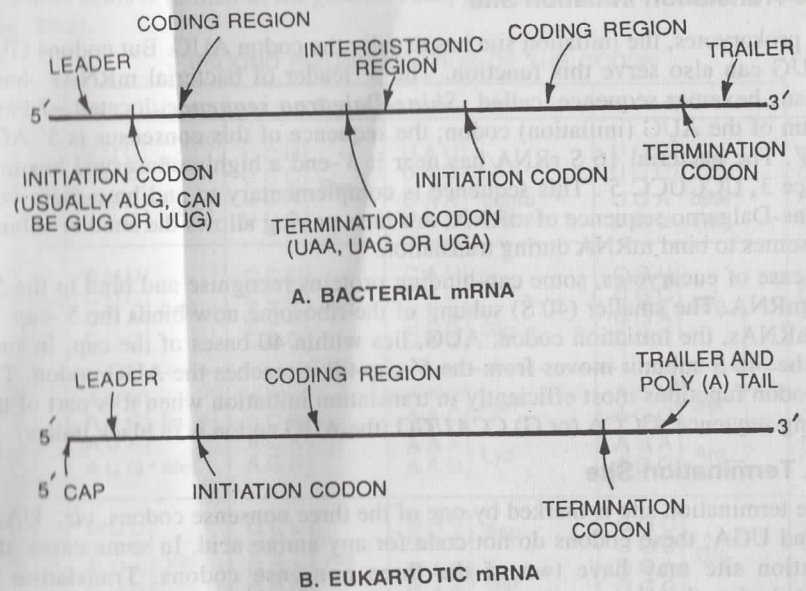


Fig. 30.1. Schematic representation of (A) polycistronic bacterial mRNA and (B) monocistronic eukaryotic mRNA.

of each polycistronic mRNAs, two or more coding regions are present; two neighbouring coding regions are separated by ***intercistronic regions*** of ~30 bases (in case of bacteria) that are not translated.

When the intercistronic region is small, ribosomes may pass directly from the previous coding region to the next coding region. But when the intercistronic region is rather long, the ribosome leaves mRNA at the termination codon of the previous coding region; a fresh initiation occurs at the beginning of the next coding region.

30.2.2. Eukaryotic mRNAs

All eukaryotic mRNAs have (1) a ***5′- cap*** (Section 29.12, Chapter 29), which is necessary for their efficient translation. This is followed by (2) an untranslated ***5′-leader*** of usually <100 bases, (3) the single ***coding region,*** and (4) an untranslated ***3′ trailer*** that may be upto 1,000 bases. In bulk (~70% of the total mRNAs) of eukaryotic mRNAs, (5) a ***poly-(A)*** tail of ~200 residues is added after transcription (Section 29.11, Chapter 29); such mRNAs are called ***poly-(A)$^+$ mRNAs.*** But about 30% of the mRNAs do not have a poly-(A) tail; they are usually termed as ***poly-(A) mRNAs.***

Eukaryotic mRNAs seem to be present in two forms, *viz.*, (1) active and (2) inactive forms. The ***active form*** is present as polysomes, and it actively supports translation. This form dissociates as ~18 S ribonucleoprotein (RNP) particles after eukaryotic polysomes are treated with EDTA (to remove the ribosomes and the newly synthesized polypeptide chains). But the ***inactive form*** does not support translation, is present as RNP particles, and separates as 200 S particles when eukaryotic polysomes are treated with EDTA. This from of mRNA is stored for later use.

30.2.3. Translation Initiation Site

In prokaryotes, the initiation site has usually the codon AUG. But codons GUG and UUG can also serve this function. The 5′ leader of bacterial mRNAs has a consensus hexamer sequence, called ***Shine-Dalgarno sequence***, located ~7 bases upstream of the AUG (initiation) codon; the sequence of this consensus is 5′ AGG AGG 3′. The bacterial 16 S rRNA has near it 3′-end a highly conserved hexamer sequence 3′ UCCUCC 5′. This sequence is complementary to and base-pairs with the Shine-Dalgarno sequence of mRNA; this base pairing allows the smaller subunit of ribosomes to bind mRNA during translation.

In case of eucaryotes, some cap-binding proteins recognise and bind to the 5′-cap of mRNA. The smaller (40 S) subunit of the ribosome now binds the 5′-cap. In some mRNAs, the initiation codon, AUG, lies within 40 bases of the cap. In such cases, the 40 S subunit moves from the 5′-cap till it reaches the AUG codon. The AUG codon functions most efficiently in translation initiation when it is part of the following sequence: GCCA (or G) CC***AUG***G (the AUG codon is in black italics).

30.2.4. Termination Site

The termination site is marked by one of the three nonsense codons, *viz.*, UAA, UAG and UGA; these codons do not code for any amino acid. In some cases, the termination site may have two of the three nonsense codons. Translation is terminated when the ribosome reaches the termination site.

Genetic Code

- Three letter code; codon.
- Non-overlapping.
- Unambiguous (some ambiguity *in vitro*).
- Universal; some protozoa and mitochondria show variation: usually, initiation and termination codons.
- 61 codons encode amino acids:
  1. 2 amino acids – one codon each.
  2. 9 amino acids – 2 codons each.
  3. 1 amino acid – 3 codons.
  4. 5 amino acids – 4 codons each.
  5. 3 amino acids – 6 codons each.
- 3 codons (*UAA, UGA, UAG*) are nonsense, *i.e.*, encode no amino acid.

30.3. THE GENETIC CODE

The number and the sequence of bases in mRNA specifying an amino acid is known as ***codon.*** Similarly, the set of bases in a tRNA that base-pairs with a codon of an mRNA is known as ***anticodon***. *The sequence of bases in an anticodon is exactly the opposite of and complementary to that present in the codon.* For example, the codon 5′ AUG 3′ has the anticodon 3′ UAC 5′, which will read CAU if read from the 5′-end as is the case for the codon (AUG). Therefore, it is customary to write the base sequence of anticodon from 5′ to 3′ direction, and to place an arrow pointing to the left to indicate the direction in which it should be read, *e.g.*, $\overleftarrow{\text{CAU}}$ (should be read as UAC; it is the anticodon for AUG). The set of all the codons that specify the

20 amino acids is termed as the ***genetic code, genetic language*** or ***coding dictionary*** (Fig. 30.2).

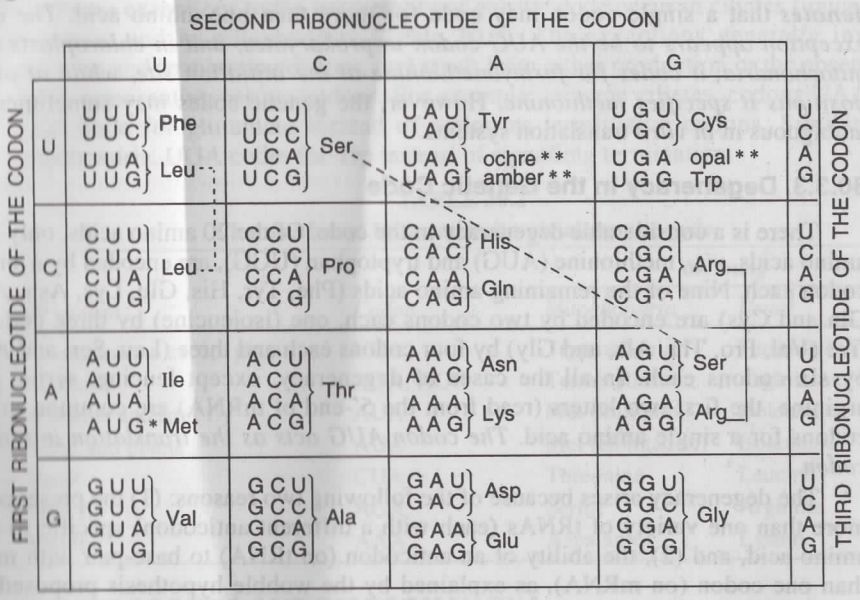


Fig. 30.2. Amino acids encoded by the 64 possible codons of the triplet code. Note that AUG (*) serves as the initiator codon and UAA (ochre), UAG (amber) and UGA (opal) (**) are nonsense codons and serve as polypeptide termination codons (this is difinitely known for UAA and UAG).

In 1954, George Gammow suggested that a codon had 3 bases. But he also suggested that codons were overlapping. In an ***overlapping coding system***, the same base of an mRNA occurs in two or 3 different codons. But in a ***nonoverlapping codon system***, each base of an mRNA occurs in only one codon, *i.e.*, is not shared by two or more codons.

In 1961, Crick and coworkers concluded from genetic studies in T_4 phage of *E. coli* that a codon had most likely three bases, and that the code was nonoverlapping, degenerate and commaless. A ***commaless*** code means that all the bases in a polynucleotide are part of codons, and that no base serves as a punctuation mark. In a ***degenerate coding system***, two or more different codons specify a single amino acid.

30.3.1. The Coding Dictionary

It became clear from the studies of Nirenberg, Ochoa, Khorana and their coworkers that a *sequence of three nucleptides in the mRNA codes for one amino acid* (a ***triplet code*** or ***codon***), *and that the code is nonoverlapping and commaless.* The meaning of all the possible 64 triplet codons has been determined beyond much doubt (Fig. 30.2). Three codons, *viz.*, UAA, UAG and UGA, are ***nonsense condons*** as they do not code for any amino acid. *These codons act as polypeptide termintion condons.*

30.3.2. Ambiguity in the Genetic Code

There is no evidence that the genetic code is ambiguous *in vivo*. ***Ambiguity denotes*** that a single codon may code for more than one amino acid. *The only exception appears to be the AUG codon in prokaryotes, and in chloroplasts and mitochondria; it codes for formylmethionine at the initiation site, while at other positions it specifies methionine.* However, the genetic codes may sometimes be ambiguous in *in vitro* translation systems.

30.3.3. Degeneracy in the Genetic Code

There is a considerable degeneracy in the code. Of the 20 amino acids, only two amino acids, *viz.*, methionine (AUG) and tryptophan (UGG), are encoded by a single codon each. Nine of the remaining amino acids (Phe, Tyr, His, Glu, Lys, Asn, Asp, Gln and Cys) are encoded by two codons each, one (isoleucine) by three codons, five (Val, Pro, Thr, Ala, and Gly) by four codons each and three (Leu, Ser, and Arg) by six codons each. In all the cases of degeneracy, except leucine, serine and arginine, the first two letters (read from the 5′-end of mRNA) are common in the codons for a single amino acid. *The codon AUG acts as the translation initiation codon.*

The degeneracy arises because of the following two reasons: (1) the presence of more than one variety of tRNAs (each with a different anticodon) specific to one amino acid, and (2), the ability of an anticodon (on tRNA) to base-pair with more than one codon (on mRNA), as explained by the wobble hypothesis proposed by Crick in 1966.

According to ***wobble hypothesis***, the first two bases of a codon pair strictly according to the normal base-pairing rules with the last two bases of the anticodon. But base-pairing at the third base of codons with the first base of anticodon is not precise (Table 30.1). An U in the first position of an anticodon may base-pair with either an A or G in the third position of codon, G in anticodon may pair with *C* or *U* in the codon, while I (inosinic acid) may pair with *U*, *C* or A in the codon.

TABLE 30.1
The pairing behaviour at the third base of codon with the first base of the anticodon (which corresponds to the third base of the codon) according to the wobble hypothesis

Base in the first position of the anticodon	*Bases recognised in the third position of the codon*
U (usually modified)	A or G
C	G only
A (always modified; usually to I)	U only
G	C or U
I	U C or A
Queuosine (Q) and its derivatives	U or C (less readily)
2-Thiouridine	Only A (does not pair with G)

30.3.4. Universality of the Genetic Code

There is considerable evidence that the *genetic code is essentially* ***universal****, that is, a codon codes for the same amino acid in all the organisms.* But there is some variation in the preference for different codons in different organisms. For example, codon AUA is used for coding isoleucine in eukaryotes. But *E. coli* does not use AUA; instead it uses AUU and AUC.

Some exceptions to the universality of genetic code occur in ciliates (unicellular protozoa) and in mitochondria (Table 30.2). The exceptions generally involve initiation and termination codons, and result from either production or the absence of tRNA representing certain codons. For example, in some ciliates, codons UAA and UAG code for glutamate instead of acting as termination codons. Similarly in mitochondria, UGA codes for Trp instead of signalling termination.

TABLE 30.2
Changes found in the mitochondrial genetic code

Organism	*Codon*	*Meaning*	
		Mitochondria	*Usual*
All organisms except plants	UGA	Tryptophan	Termination
Mammal	AGA, AGG	Termination	Argininine
Mammal and fruit fly	AUA	Met (initiation)	Isoleucine
Yeast and plants	AUA	Met (elongation)	Isoleucine
Yeast	CUA	Threonine	Leucine
Fruit fly	AGA	Serine	Arginine
Molds	AAA	Asparagine	Lysine

30.4. RIBOSOMES AND RIBOSOMAL RNA

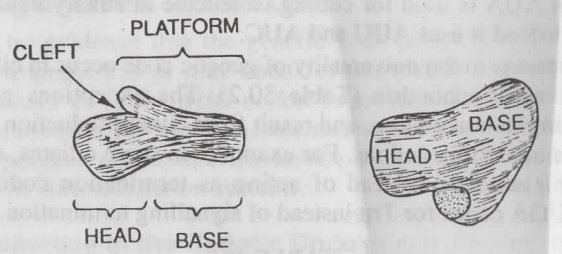
Ribosomal RNA (rRNA) occurs in association with proteins and is organised into almost spherical bodies of about 200 Å diameter called ***ribosomes***. Ribosomes of prokaryotes are of 70 S size and are composed of 60% rRNA and 40% protein. They dissociate into a smaller 30 S subunit and a larger 50 S subunit. The cytoplasmic ribosomes of eukaryotes are of the size 80 S and contain 40% rRNA and 60% protein. They consist of a 40 S and a 60 S subunit.

The 30 S subunit of prokaryotic ribosomes has a single 16 S rRNA molecule (about 1,541 nucleotides), which is associated with 21 different 'S' proteins. The larger 50 S subunit has a 23 S rRNA (about 2,904 nucleotides) and a 5 S rRNA (120 bases) molecule complexed with 31 different 'L' proteins. The basic structure of ribosomes is due to their rRNA, and the proteins play a secondary role.

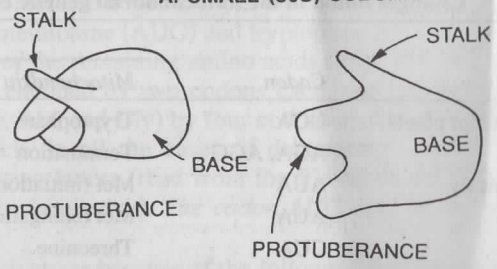
In the case of eukaryotic ribosomes, the 40 S subunit has a single 18 S rRNA molecule (1,874 bases) and 33 different proteins. The larger 60 S subunit has one molecule each of 28 S (4,718 bases), 5.8 S (160 bases) and 5 S (120 bases) rRNA and 49 different proteins.

30.4.1. Ribosome Organisation

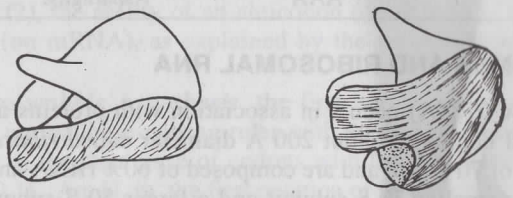
Ribosomes have an asymmetrical organisation. The 30 S subunit of prokaryotic ribosomes has an asymmetrical and elongated shape. It has a constricted region and a cleft that separates the head from the base, while its platform projects out of the base (Fig. 30.3A). The larger 50 S subunit has a fairly compact body from which a central protuberance and a stalk stick out (Fig. 30.3B). The platform of the small subunit fits into the large subunit, and it is likely that a tunnel is formed between the two subunits.



A. SMALL (30S) SUBUNIT



B. LARGE (50S) SUBUNIT



C. 70S RIBOSOME

Fig. 30.3. Schematic representation of the organisation of prokaryotic ribosomes.

Ribosomes

- Asymmetrical bodies of 200 Å dia.
- *Prokaryotic*: 70 S; 30 S and 50 S subunits; rRNA 16 S (30 S subunit), and 23 S and 5 S (50 S subunit).
- *Eukaryotic*: 80 S; 40 S and 60 S subunits; rRNA; 18 S (40 S subunit), and 28 S, 5.8 S and 5 S (60 S subunit).
- *Functional Sites:* In *E. coli* ribosome:
  1. mRNA site.
  2. Peptidyl transferase site.
  3. EF- Tu site.
  4. EF-G site.
  5. 5 S RNA site.
  6. Exit site.

Each ribosomal protein seems to bind to specific sites on the rRNA. Binding of some proteins seems to generate the binding sites for other proteins. Ribosome subunit assembly seems to take place in a series of steps: a group of proteins associates with rRNA, and the structure then folds so that the next group of proteins can join.

30.4.2. Functional Sites on Ribosomes

The two subunits of ribosomes dissociate in low magnetisum (Mg^{++}) concentration and reassociate in high Mg^{++} concentration. Further, *each ribosome dissociates into the two subunits after every time it finishes reading an mRNA molecule. The two subunits, reunite, at random, to produce complete ribosome once the smaller subunit attaches itself to the initiation site of an mRNA molecule.* The larger subunit of the ribosomes serves to stabilize the amino acid-bound tRNA, while the smaller subunit stabilizes the mRNA. This allows the formation of peptitide bonds between the two amino acids attached to the tRNA molecules base-paired with the two adjacent codons on mRNA.

A complete ribosome particle has two distinct and adjacent sites for the attachment of aminoacyl-tRNAs (amino acid carrying tRNA): an aminoacyl attachment site (***site A***) and a peptidyl site (***site P***). An aminoacyl-tRNA first attaches to site A, the kind of aminoacyl-tRNA being determined by the sequence of mRNA (codon) attached to the site A. The aminoacyl-tRNA, along with the mRNA codon, now moves to the site P, making the site A available for the attachment of a new aminoacyl-tRNA. In addition, the ribosome has following sites as well: (1) mRNA site, (2) peptidyl transferase site, (3) EF-Tu site, (4) EF-G site, (5) 5 S RNA site and (6) exit site. The ribosomes may be divided into two functional domains: (1) the ***translational domain*** comprises 2/3 of the ribosome and contains all the sites,

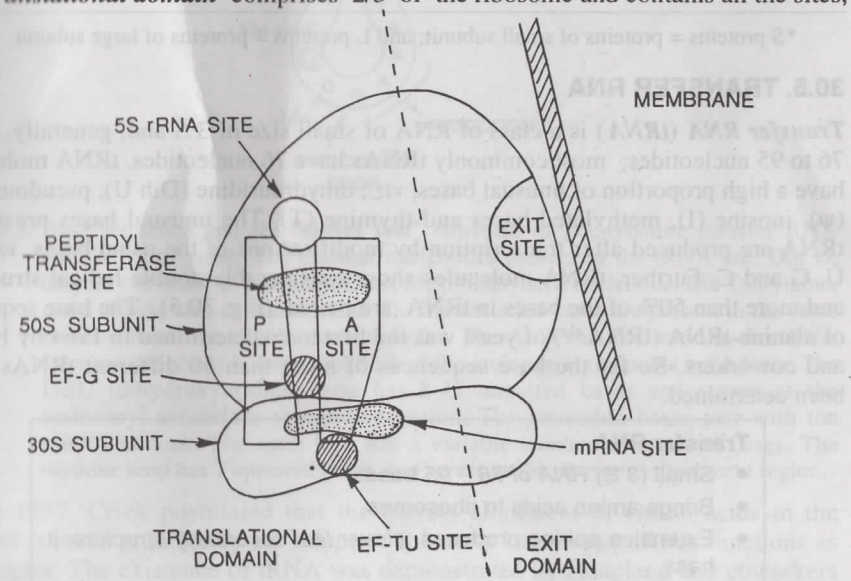


Fig. 30.4. A schematic representation of the different functional sites on ribosomes.

except the exit site, which is located in the second domain called (2) the ***exit domain***. In eukaryotes, the exit domain of ribosomes is associated with the ER.

The *A* and *P* sites are formed by both 30 S and 50 S subunits. The 30 S subunit has the mRNA and EF-Tu binding sites, while EF-G, 5 S rRNA, peptidyl transferase and exit sites are located on the larger 50 S subunit (Fig. 30.4; Table 30.3).

TABLE 30.3

A list of the *E. coli* ribosome functional sites and the ribosome components involved in the functions (for prokaryotic ribosomes)

Site	*Function*	*Location*	*Components*
mRNA binding	Binds to mRNA and IF factors	30S, near P site	3′-terminal region of 16S rRNA + 7 S proteins*
P site	Binds fMet- $tRNA_f$ and peptidyl-tRNA	Mostly 30S	Near 3′-end of 16S rRNA + 6 L proteins*
A site	Binds aminoacyl-tRNA	Mostly 50S	Both 16S and 23S rRNA + 6 L proteins*
Exit site	Binds deacylated-tRNA	50S	23S rRNA
Peptidyl transferase	Transfers peptide to aminoacyl-tRNA	50S, near P and A sites	23S rRNA + 5 L proteins
5S rRNA	Not known	50S central protuberance	3 L proteins
EF-Tu binding	Aminoacyl-tRNA entry	30S exterior	
EF-G binding	Translocation	50S near 30S	
L7/L12	Needed for GTPase activity	50S stalk	L7, L12 proteins

*S proteins = proteins of small subunit; and L proteins = proteins of large subunit.

30.5. TRANSFER RNA

***Transfer RNA* (*tRNA*)** is a class of RNA of small size of 3 S and, generally, it has 76 to 95 nucleotides; most commonly tRNAs have 76 nucleotides. tRNA molecules have a high proportion of unusual bases, *viz.*, dihydrouridine (Dih U), pseudouridine (ψ), inosine (I), methylated bases and thymine (T). The unusual bases present in tRNA are produced after transcription by modifications of the usual bases, *viz.*, A, U, G and C. Further, tRNA molecules show considerable double helical structure, and more than 50% of the bases in tRNA are paired (Fig. 30.5). The base sequence of alanine-tRNA ($tRNA^{ala}$) of yeast was the first to be determined in 1964 by Helley and coworkers. So far the base sequences of more than 50 different tRNAs have been determined.

Transfer RNA

- Small (3 S) *RNA of 76 – 95* bases.
- Brings amino acids to ribosomes.
- Extensive pairing produces, *clover leaf secondary structure;* it has:
  1. Amino acid acceptor region CCA at 3′-end.
  2. A thymine loop.
  3. Anticodon loop; always at 66 Å from 3′–end (CCA).
  4. *DHU* loop (toward 5′-end).
  5. An extra loop between thymine and anticodon loops.
- All tRNAs acquire '*L*–shaped' *tertiary structure*: CCA lies at the tip of one arm, and anticodon lies at the tip of the other.

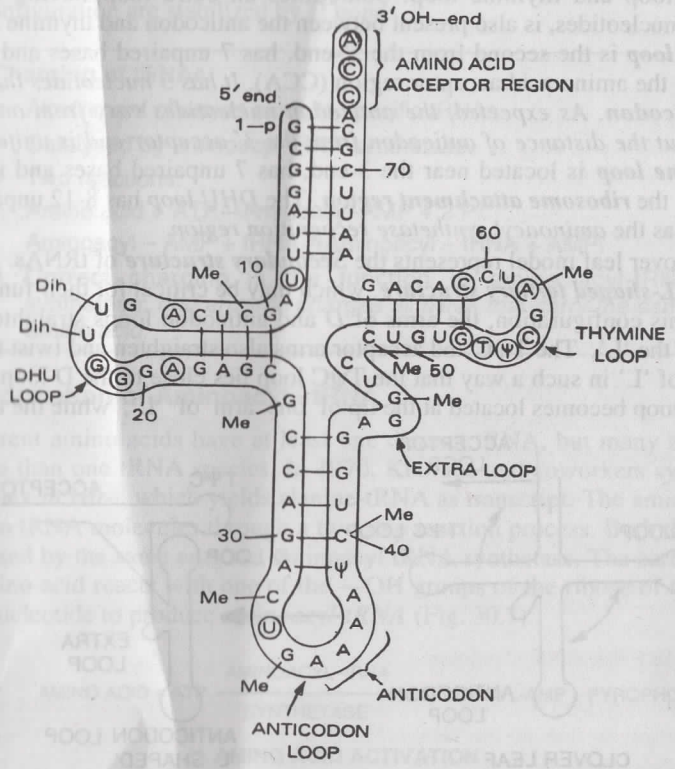


Fig. 30.5. Base sequence and the 'clover leaf' structure of phenylanine transfer RNA ($tRNA^{Phe}$) of yeast. The unusual nucleotides present in the tRNA are: Dih U, dihydrxy uridine; ψ, psueodouridine; Mex, methylated nucleotide, and T, thymine; some tRNAs also contain inosine (I). The encircled bases are present in the same position in all the tRNAs studied so far. The A: T and G:C ratio in the tRNA fractions are about 1, and about 80% of the molecule is in double helix form. The DHU (dihydroxyuridine) loop has 8-12 unpaired bases and serves as the aminoacyl synthetase recognition region. The anticodon bases pair with the mRNA codons. The extra loop has a variable number of unpaired bases. The thymine loop has 7 unpaired bases, and serves as the ribosome attachment region.

In 1957, Crick postulated that the correct alignment of amino acids in the proteins is made possible by an '*adaptor*. It is now known that tRNA functions as the adaptor. The existence of tRNA was demonstrated by Hoagland and coworkers in 1957.

30.5.1. Secondary and Testiary Structure

A ***'clover leaf model'*** of tRNA structure has been proposed to account for its structural and functional properties (Fig. 30.5). Nucleotides of all the tRNAs sequenced so far can be satisfactorily arranged according to this model. The three nucleotides at the 5′-end are CCA in all the types of tRNA molecules. This segment is the ***amino acid acceptor region.*** *The amino acid binds to one of the OH groups of the ribose of the terminal adenine.* The tRNA molecule has three loops: DHU loop,

anticodon loop and thymine loop. Sometimes an extra loop, having a variable number of nucleotides, is also present between the anticodon and thymine loops. The ***anticodon loop*** is the second from the 5′-end, has 7 unpaired bases and is located opposite to the amino acid acceptor region (CCA). *It has 3 nucleotides that function as the* ***anticodon****. As expected, the anticodon nucleotides vary from one tRNA to another. But the distance of anticodon from the 3′ acceptor end is uniformly 66Å.* The ***thymine loop*** is located near the 3-end, has 7 unpaired bases and is likely to function as the ***ribosome attachment region***. The *DHU loop* has 8-12 unpaired bases and serves as the *aminoacyl synthetase recognition region.*

The clover leaf model represents the ***Secondary structure*** of tRNAs. All tRNAs acquire an ***L-shaped tertiary structure***, which may be critical for their function (Fig. 30.6). In this configuration, the arms of *D* and anticodon loops straighten to form one arm of the 'L'. The TψC and acceptor arms also straighten and twist to form the other arm of 'L' in such a way that the TψC loop lies close to the D loop. Thus, the anticodon loop becomes located at the tip of one arm of 'L', while the amino acid

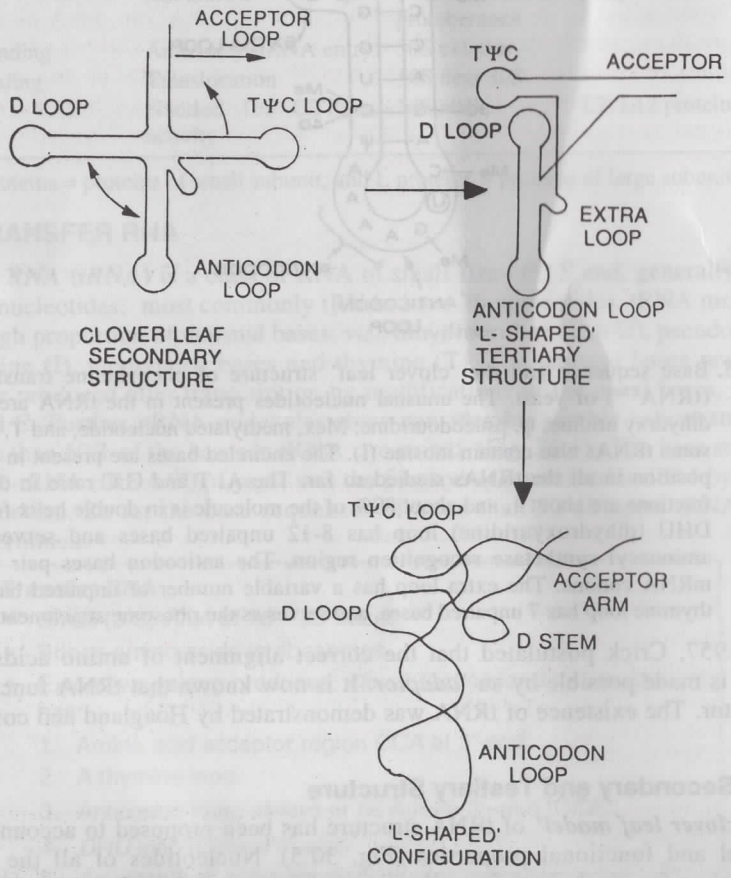


Fig. 30.6. The tertiary configuration (L-shaped) of the clover-leaf secondary structure of tRNA molecules.TψC loop = thymine loop; D loop = DHU loop.

acceptor region is present at the tip of the other arm. Both clover leaf and the 'L'-shaped configuration are stabilized by H bonds.

Charging of tRNAs

- Attachment of amino acids to specific tRNAs.
- Catalyzed by aminoacyl-tRNA synthetase.
- Two reactions:
  Amino acid + ATP→Aminoacyl-AMP + 2 P*i*.
  Aminoacyl – AMP + tRNA→Aminoacyl – tRNA + AMP.
- Correct charging is the function of amino acyl-tRNA synthetase; *one enzyme charges all tRNA that carry the same amino acid.*

30.5.2. Formation of Aminoacyl - tRNA

Different amino acids have at least one distinct tRNA, but many amino acids have more than one tRNA species. In 1970, Khorana and coworkers synthesized a DNA duplex *in vitro*, which yields alanine-tRNA as transcript. The amino acids are attached to tRNA molecules through a two step reaction process. Both the reactions are catalysed by the same enzyme, aminoacyl tRNA synthetase. The carboxyl group of the amino acid reacts with one of the —OH groups of the ribose of the terminal adenine nucleotide to produce ***aminoacyl-tRNA*** (Fig. 30.7).

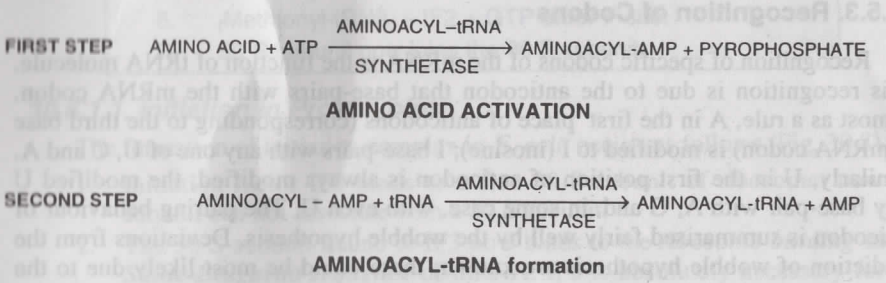


Fig. 30.7. Steps in the attachment of amino acids to tRNA. Aminoacyl-AMP remains complexed with the activating enzyme, aminoacyl-tRNA synthetase.

1. *First Step.* It consist of amino acid activation, in which the amino acid molecule reacts with and ATP molecule to yield an aminoacyl -AMP (*aminoacyl adenylate*) molecule (Fig. 30.7).
2. *Second Step.* The amino acid from aminoacyl–AMP molecule is then transferred to a tRNA molecule that is specific for the amino acid, and AMP is released.

The aminoacyl-AMP complex remains tightly bound to the enzyme aminoacyl–tRNA synthetase during the entire reaction, *i.e.,* both the steps. Each amino acid has its own distinct aminoacyl - tRNA synthetase, but some amino acids may have more than one activating enzyme. The attachment of an amino acid to the concerned tRNA to yield aminoacyl-tRNA is called ***charging of tRNA***. A single amino-acyl-

tRNA synthetase enzyme charges all the different tRNAs that carry the same amino acid, *i.e.*, have anticodons specifying the same amino acid. However, errors do occur during charging of tRNAs, the rate being quite low, *e.g.*, 1.5×10^{-5}.

The correct alignment of the aminoacyly tRNA on specific codons of mRNA (associated with a ribosome) is made possible by the specificity of tRNA molecules (presence of anticodons). However, recognition of the specific tRNA to which a particular amino acid is to be attached is accomplished by the amonoacyl synthetase specific for that amino acid.

Codon Recognition

- Function of anticodon; written as CAU (5′– CAU –3′).
- In the first position of anticodon :
  - (*i*) A is modified to I; I base-pairs with U, C or A.
  - (*ii*) U is always modified; U may base-pair with A or G.
  - (*iii*) G and C base-pair correctly.
- Pairing at second and third places of anticodon is always correct
- Above behaviour is explained by *wobble hypothesis* (due to Crick)

30.5.3. Recognition of Codons

Recognition of specific codons of the mRNA is the function of tRNA molecule. This recognition is due to the anticodon that base-pairs with the mRNA codon. Almost as a rule, A in the first place of anticodons (corresponding to the third base of mRNA codon) is modified to I (inosine); I base-pairs with any one of U, C and A. Similarly, U in the first position of anticodon is always modified; the modified U may base-pair with A, G and, in some case, with even U. The pairing behaviour of anticodon is summarised fairly well by the wobble hypothesis. Deviations from the prediction of wobble hypothesis are known; these could be most likely due to the effect of bases in the neighbourhood of the concerned anticodons. According to wobble hypothesis, a total of 32 tRNA are required to recognise all the 61 codons that encode amino acids (3 codons are nonsense). Ribosomes ordinarily move 3 bases at a time along the mRNA, this is called ***translocation.*** Regular translocation is based on the geometry of tRNA/mRNA interaction, rather than on the actual counting of the mRNA bases.

Translation Process

- Mediated by ribosomes.
- *Three Steps:*
  1. Initiation
  2. Elongation
  3. Termination.

30.6. THE TRANSLATION PROCESS

Translation begins near the 5′-end of mRNA and progresses toward its 3′–end. In terms of the polypeptide being synthesized, it starts with the N-terminal end of the polypeptide, while the C-terminal amino acid is the last to be added to the chain. The process of translation may be divided into the following 3 steps: (1) initiation, (2) elongation and (3) termination. These events are briefly described separately for prokaryotes and eukaryotes.

30.6.1. Initiation

Initiation comprises all the events that precede the formation of the first peptide bond. It includes the following events: (1) binding of the smaller subunit of ribosome to mRNA and (2) binding of the first or initiator aminoacyl-tRNA to the *P* site of ribosome; this completes the formation of ***initiation complex***. *Initiation is a relatively slow step in translation.*

Initiation: Prokaryotes

- All events occurring prior to the formation of first peptide bond.
- *The sequence of events:*
  1. IF3 binds 30 S subunit.
  2. IF3-30 S complex binds Shine-Dalgarno sequence.
  3. IF1 now joins 30 S subunit.
  4. 30 S subunit positions its *P* site at AUG codon.
  5. $_f$Methionyl- tRNA$_f$ binds IF2.
  6. $_f$Methionyl-tRNA$_f$ – IF2 + GTP enter *P* site.
  7. 50 S subunit now joins the 30 S subunit.

30.6.1.1. Initiation in Prokaryotes

The formation of initiation complex in *E. coli* occurs as follows (Fig. 30.8).

1. Initiation factor IF3 binds to a free 30 S subunit of ribosome; now this subunit can not associate with a 50 S subunit.
2. The 30 S subunit bound to IF3 now binds to the ***ribosome binding site*** or ***Shine-Dalgarno sequence*** of mRNA. IF3 is absolutely necessary for this binding.
3. IF1 binds to 30 S subunit bound to mRNA and possibly stabilizes the initiation complex.
4. The 30 S-IF3 complex ultimately positions itself on the mRNA in such a way that the initiation codon AUG is located in the P (and not the A) site of the 30 S subunit.
5. The formylmethionyl-tRNA$_f$ (fmet - tRNA$_f$) binds to IF2 to form a binary complex. The binary complex along with one molecule of GTP binds the codon AUG in the *P* site. Only fmet–tRNA$_f$ can enter the P site; all other tNAs can not gain a direct access to the P site.
6. The 50 S subunit of ribosome now joins the 30 S subunit; GTP is hydrolysed, and IF2 and IF3 are released. The ***initiation complex*** is now complete with an active ribosome in which A site is vacant, while the P site is occupied by fmet - tRNA$_f$.

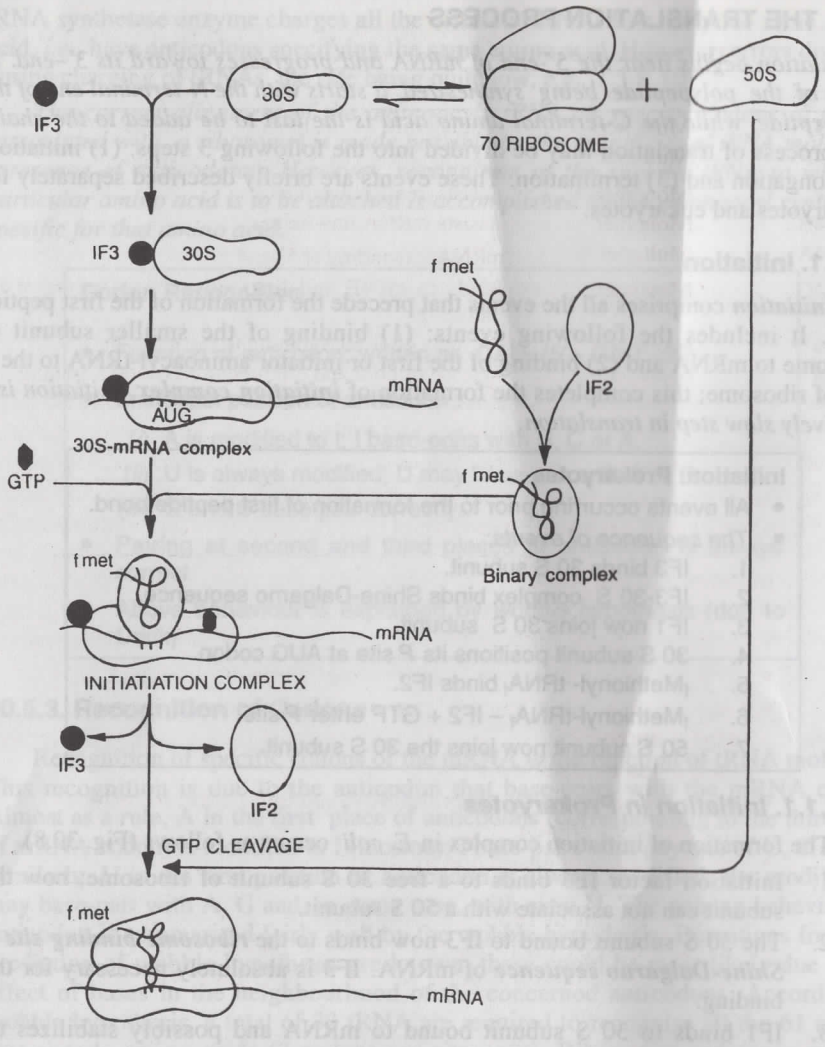


Fig. 30.8. Initiation of translation in prokaryotes.

30.6.1.2. Initiation in Eukaryotes

Initiation in eukaryotes requires at least nine different factors (Table 30.4); it proceeds as follows.

1. First of all, initiation factor eIF4 (e prefix denotes eukaryotic) binds the 5′-cap of mRNA. This allows eIF4A and eIF4B to unwind any secondary structure that may be present in the first 15 bases of mRNA.
2. The initiator aminoacyl-tRNA, *viz.*, Met-tRNA$_i$ (methionyl – tRNA$_i$) binds to eIF2-GTP complex.
3. The Met-tRNA$_i$-eIF2-GTP complex now binds in the P site of a free 40 S subunit of the ribosome.

TABLE 30.4
The translation initiation factors of mammalian reticulocytes

Factor	*Structure*	*Functions*
eIF3*	Multimer	Binding mRNA; maintains free 40 S subunits
eIF4F	Multimer**	Binding mRNA at 5′-cap and unwinding
eIF1	Monomer	Assists in mRNA binding
eIF4B	Monomer	Assists mRNA unwinding
eIF4A	Multimer	mRNA unwinding at 5′-end
eIF4G	Monomer	Binds to eIF4E and links other components of the multimer
eIF6	Monomer	Prevents 40 S-60 S joining in the free state
eIF5	Monomer	Release of eIF2 and eIF3 required for 60 S joining
eIF4C	Monomer	Binding of 60 S to 40 S associated with mRNA
eIF2	Trimer	Binding Met-tRNA$_i$,
eIF4D	Monomer	Not known

*e, This prefix denotes that the factors are of eukaryotic origin.
**, Includes eIF4E, eIF4A, eIF4G and Mnk1.

Initiation: Eukaryotes
- eIF4 binds 5′-cap of mRNA.
- Met-tRNA$_i$ binds eIF2-GTP complex.
- The above complex then binds P site of a free 40 S subunit.
- This 40 S subunit binds 5′-cap of mRNA (assisted by eIF3, eIF4 A and eIF4B.
- eIF6 binds a 60*s* subunit.
- eIF6 – 60 S complex binds the 40 S subunit bound to mRNA (mediated by eIF4C).
- eIF2 and eIF3 released due to eIF5.

4. This 50 S subunit of ribosome now binds to the 5′-cap of mRNA; it then moves along RNA till it reaches the initiation codon AUG. This process required eIF3 and several other factors like eIF4, eIF4 A and eIF4B (Fig. 30.9).
5. Factor eIF6 binds to the 60 S subunit and prevents its association with a free 40 S subunit. Now 60 S-eIF6 complex joins the 40 S subunit bound to mRNA; both eIF2 and eIF3 are released due to the action of eIF5. The joining of 40 S and 60 S subunits is mediated by eIF4 C.
6. Once the complete 80 S ribosome is assembled at the AUG codon, probably all the remaining initiation factors are released.

30.6.2. Elongation

Elongation includes all the reactions from the formation of the first peptide bond to that of the last peptide bond of the polypeptide chain. Amino acids are added one at a time to the growing peptide chain. It is the most rapid step of translation: in bacteria, it proceeds at the rate of ~15 amino acids/second at 37°C, while in eukaryotes its rate is ~2 amino acids/second. The elongation step in bacteria is briefly described below.

The complete ***initiation complex*** has the following features: (1) ribosome is complete and ready for translation, (2) *P* site of ribosome is occupied by the initiator aminoacyl-tRNA (fmet-tRNA$_f$) and (3) the A site is free. The P site of ribosome must be occupied by peptidyl-tRNA (tRNA carrying a peptide or fmet-tRNA for the A site to be in a proper conformation to allow entry to an amino-acyl-tRNA. Any amino-acyl-tRNA (except fmet-tRNA$_f$) can enter the free ***A site of an active ribosome***, depending mainly on the codon of mRNA present at the A site.

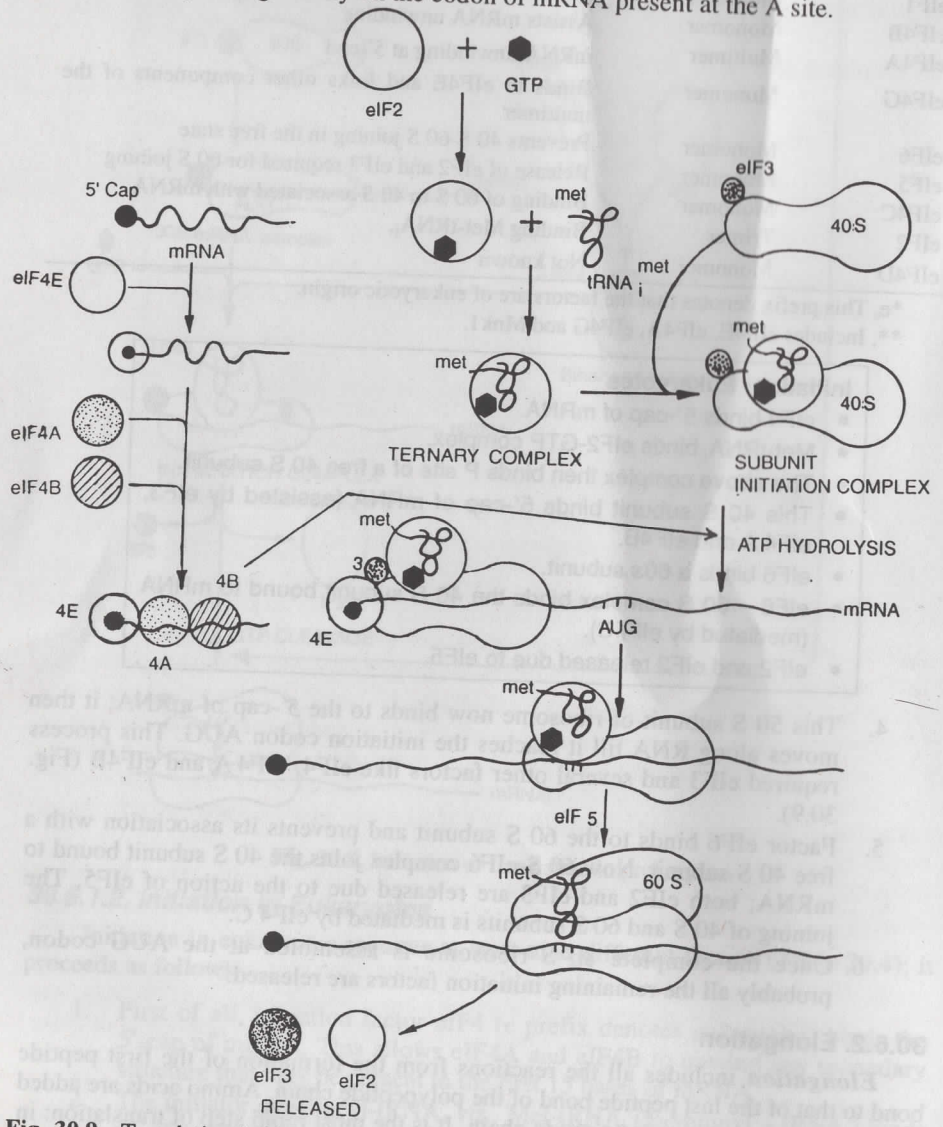


Fig. 30.9. Translation initiation in eukaryotes. For simplicity, only some of the initiation factors are depicted.

The entry of an aminoacyl-tRNA into a free A site is mediated by the elongation factor EF-Tu. Active EF-Tu is bound to GTP. EF-Tu-GTP complex binds to an aminoacyl-tRNA to form a ***ternary complex.*** This ternary complex now binds to the free A site. As stated before, the choice of particular aminoacyl -tRNA (present as ternary complex) entering the A site depends on the mRNA codon present at the A site. Now GTP is cleaved and EF-Tu-GDP complex is released; this is essential for peptide bond formation.

EF-Tu-GDP complex is inactive. It is activated by factor EF-Ts, which first of all displaces the GDP from EF-Tu-GDP complex to give rise to EF-Tu. EF-Ts complex. Now EF-Ts is itself displaced by GTP to yield the active EF-Tu-GTP complex (Fig. 30.10). In this manner, EF-Tu circulates between active and inactive states.

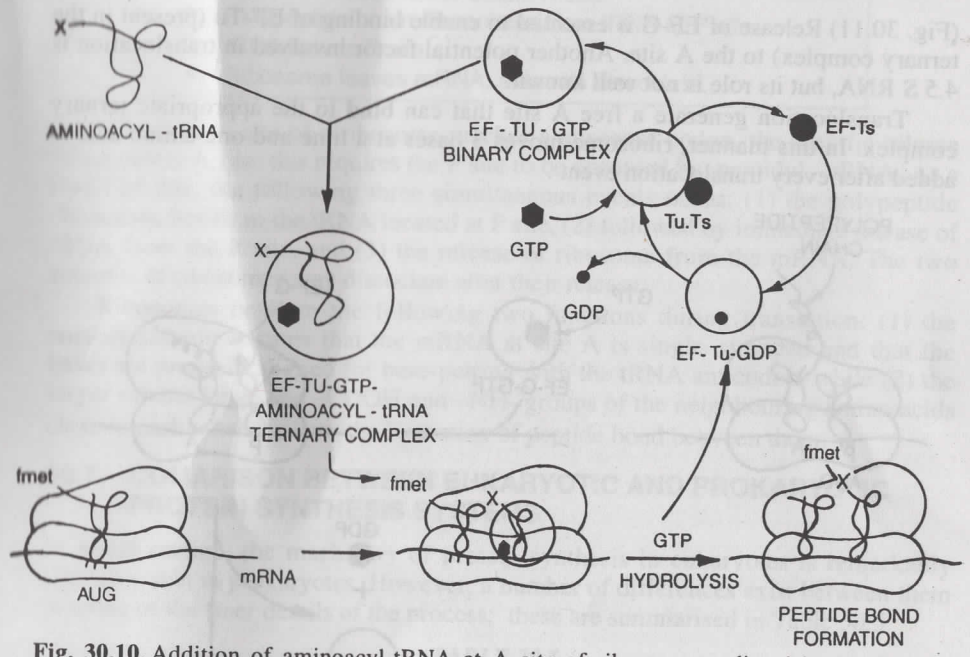


Fig. 30.10. Addition of aminoacyl-tRNA at A site of ribosome mediated by EF-Tu (in bacteria)

Entry of the ternary complex into the A site aligns the amino acids attached to tRNAs present at P and A sites in a proper manner for peptide bond formation. Peptide bond formation is catalyzed by ***peptide transferase*** activity of the 50 S subunit of ribosome (activityof the 23 S rRNA of the 50 S subunit).

After peptide bond formation, the P site has a free tRNA, while the A site has a peptidyl –tRNA. Now the ribosome moves 3 bases along the mRNA; this removes the free tRNA from the P site, and moves the peptidyl tRNA from A site into the now free P site. In prokaryotes, the free tRNA leaves the ribosome via the exist site, while in eukaryotes it directly goes into the cytosol. This whole process if called ***translocation***; it requires elongation factor EF-G and GTP. EF-G-GTP complex binds to the EF-G site of 50 S subunit, GTP is hydrolysed, the ribosome moves 3 bases along the mRNA and EF–G and GDP are released from the ribosome

Elongation

- Rate: in bacteria, ~ 15 amino acid/sec at 37°C; in eukaryotes, ~2 amino acids/sec
- *Elongation process:* In *E. coli*,
  1. EF-Tu·GTP complex binds an amino acyl–tRNA.
  2. Appropriate EF-Tu·GTP-aminoacyl-tRNA enters A site.
  3. Peptidyl transferase site catalyzes peptide bond formation.
  4. Free tRNA leaves P site via exit site.
  5. Ribosome moves 3 bases along mRNA (*translocation*); mediated by EF-G·GTP complex.
  6. The A site now becomes free; steps 1-5 repeated.

(Fig. 30.11) Release of EF-G is essential to enable binding of EF-Tu (present in the ternary complex) to the A site. Another potential factor involved in translocation is 4.5 S RNA, but its role is not well known.

Translocation generate a free A site that can bind to the appropriate ternary complex. In this manner, ribosome moves 3 bases at a time and one amino acid is added after every translocation event .

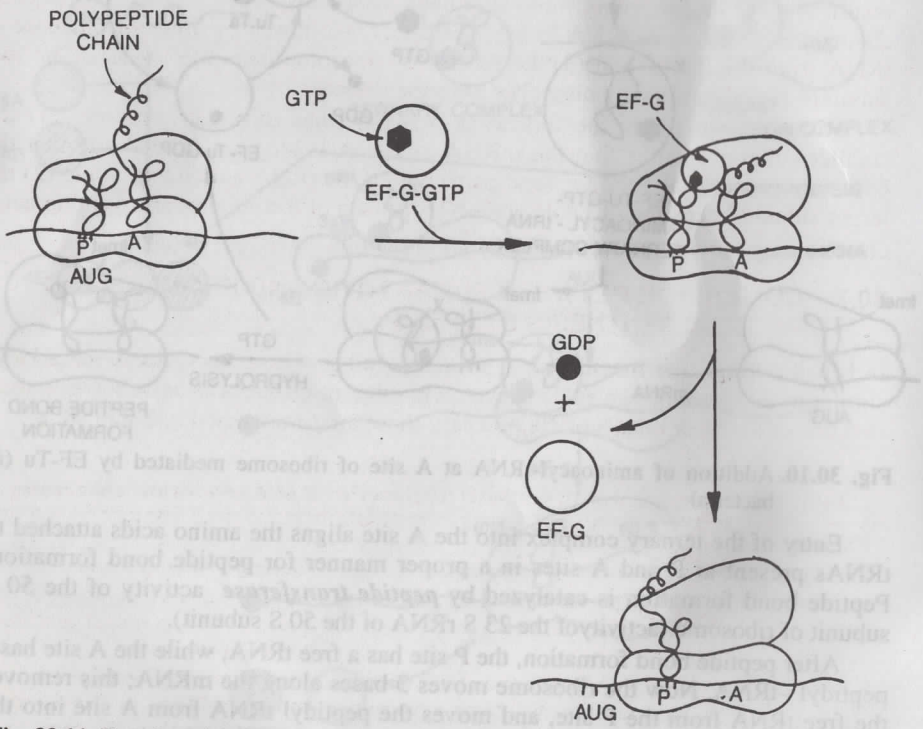


Fig. 30.11. Translocation mediated by EF-G. P and A denote the P and A sites of ribosome.

30.6.3. Termination

Termination describes those events that occur after the formation of the last peptide bond of the polypeptide; they culminate in the separation of the ribosome

from mRNA. Termination is signalled by the nonsense codons UAG, UAA and UGA, and is mediated by certain proteins called ***release factors***. *E. coli* has two release factors, RF1 and RF2. Factor RF1 recognizes UAA and UAG, while RF2 recognizes UGA and UAA.

Termination

- Events that occur after formation of the last peptide bond.
- Mediated by release factors; *E. Coli* has two factors:
  RF1 acts with UAA, UAG.
  RF2 acts with UGA and UAA.
- When A site reaches a nonsense codon:
  1. Specific release factor enters A site.
  2. Polypeptide chain dissociates from tRNA at P site.
  3. tRNA leaves P site.
  4. Ribosome leaves mRNA; subunits dissociate.

When the A site of a ribosome reaches a nonsense codon, the specific release factor enters A site; this requires the P site to be occupied by peptidyl – tRNA. As a result of this, the following three simultaneous events occur: (1) the polypeptide chain detaches from the tRNA located at P site, (2) followed by immediate release of tRNA from the P site, and (3) the release of ribosome from the mRNA. The two subunits of ribosomes may dissociate after their release.

Ribosomes perform the following two functions during translation: (1) the smaller subunit ensures that the mRNA at site A is single–stranded and that the bases are properly aligned for base-pairing with the tRNA anticodon, while (2) the larger subunit keeps the –COOH and $-NH_2$ groups of the neighbouring amino acids close to each other to allow the formation of peptide bond between them.

30.7. COMARISON BETWEEN EUKARYOTIC AND PROKARYOTIC PROTEIN SYNTHESIS SYSTEMS

In broad outline, the machinery of protein synthesis in eukaryotes is remarkably similar to that in prokaryotes. However, a number of differences exist between them in terms of the finer details of the process; these are summarised in Table 30.5.

TABLE 30.5

Differences between protein synthesis machineries of prokaryotes and eukaryotes

Characteristic	*Prokaryotes*	*Eukaryotes*
mRNA	Generally polycistronic.	Generally, monocistronic.
	In general, there is no processing after transcription.	Mature mRNA molecules are processed from longer precursor hnRNA molecules.
	Very short half-life of 2-5 minutes; degradation may begin even before transcription is complete.	Much longer half-life of ~6 hr or more; mRNA in seeds may survive more than 100 years.
	Poly-(A) tail not found.	Most mRNA's have 100-200 adenine nucleotides at the 3′-end, producing a poly-(A) tail.
	5′-Cap not present.	Partially methylated 5′-cap present.

	Shine-Dalgarno sequence present at the 5′-end	This sequence is absent.
rRNA and ribosomes	Ribosomes of 70 S; dissociate into 30 S and 50 S subunits.	Ribosomes of 80 S; dissociate into 40 S and 60 S subunits.
	Ribosomes free in the cytoplasm.	Ribosomes attached to endoplasmic reticulum as well as free in the cytoplasm.
	rRNA of 16 S (30 S subunit), 23 S and 5 S (50 S subunit).	rRNA is 18 S (40 S subunit), 28 S, 5.8 S and 5 S (60 S subunit)
tRNA	$tRNA_f^{met}$ carries formylmethionine to the initiation site of mRNA	$tRNA_i^{met}$ carries methionine to the initiation site of mRNA.
Translation	Translation often simultaneous with transcription.	Translation is not simultaneous with transcription.
	Translation occurs while mRNA is attached to the chromosome.	mRNA has to be processed and transported out of the nucleus as translation occurs mostly in the cytoplams.
	Translation initiation involves base-pairing between 16 S rRNA and mRNA in the Shine-Dalgarno consensus sequence.	Translation initiation is based on the recognition of the 5′-cap of mRNA by the 40 S subunit of ribosomes and by some proteins.
	Formylmethionine is incorporated by the initiation codon AUG.	Methionine is incorporated at the initiation point by the codon AUG.
	Initiation site of mRNA first binds with the 30 S subunit of ribosome, which then accepts formyl-methionyl-$tRNA_f^{Met}$ complex.	Methionyl-$tRNA_i^{Met}$ first binds to the 40 S subunit, to which an mRNA molecule is then attached to form the initiation complex.
Coding dictionary	Isoleucine encoded by AUU, and AUA may not be used.	AUA used for encoding isoleucine.

QUESTIONS

1. List the various components of protein synthesis. Briefly describe the mode of protein synthesis in prokaryotes.
2. Give a brief account of the functional organization and other relevant factors of mRNA. Briefly explain the manner in which ribosomes recognize the 5′-end of mRNA molecules.
3. What is genetic code ? Explain its essential features and discuss the correctness of universality of genetic code.
4. Briefly describe the structure and function of ribosomes (or tRNA).
5. Briefly describe the mechanism of translation initiation in prokaryotes and eukaryotes and list the differences in the translation processes of the two groups of organisms.
6. Write short notes on following:
   (*i*) ribosome binding site, (*ii*) ribosomes,
   (*iii*) tRNA, (*iv*) mRNA,
   (*v*) genetic code, (*vi*) wobble hypothesis,
   (*vii*) translation initiation, (*viii*) translation elongation,
   (*ix*) translation termination, (*x*) translocation (ribosome).

31

Regulation of Gene Action I. *Prokaryotes*

31.1. INTRODUCTION

Gene action or ***gene expression***, *in the classical sense*, refers to the production of the concerned trait or phenotype by a gene. The phenotype production, however, is based on production of a specific protein encoded by this gene and the subsequent participation of this protein in a metabolic/ biosynthetic pathway. The sequence of events in phenotype production may, therefore, be written as follows (Fig. 31.1).

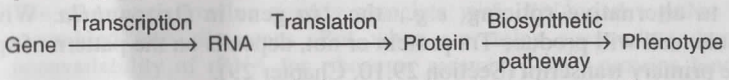


Fig. 31.1. Sequence of events in the production of the phenotype by a gene.

But in *molecular, terms,* ***gene expression*** is considered to mean transcription, *i.e.,* production of RNA transcript of the gene. *The expression of gene may be assayed in terms of RNA production, protein/enzyme activity or the specific phenotype produced.* In different tissues of eukaryotes, different sets of proteins and/or enzymes are produced as per the specific needs of the tissues. For example, RBCs contain large quantities of haemoglobin, while white blood cells (WBCs) have no haemoglobin at all. Similarly, prokaryotes require different enzymes at different points of time. Usually, prokaryotes produce only those enzymes that are needed by them at that particular time. Thus no cell produces all the proteins that it is capable of; it produces only that group of proteins that are required for its efficient function. This ability of cells is precisely regulated, the level at which this regulation occurs varies to a considerable extent.

31.2. LEVELS OF REGULATION OF GENE ACTION

The mechanism by which the expression of different genes is controlled in different tissues and/or at different times is called ***gene regulation***, ***regulation of gene expression*** or ***regulation of gene action***. The expression of a gene may be subject to regulation at one or more of the following levels: (1) gene amplification, destruction or distribution, (2) transcription, (3) post-transcription, (4) translation and (5) post-translation.

31.2.1. Gene Amplification, Destruction or Distribution

Regulation at this level determines whether a gene is present in a cell, and if present, the number of copies in which it is present. In general, those genes whose products are required in large quantities are present in multiple copies in the genone.

In addition, additional copies of some genes are also produced by replication. These remain in extrachromosomal state; the copies are usually produced at specific times, are used for transcription, and are subsequently digested, this phenomenon is called ***amplification***. An excellent example of amplification is the production of extrachromosomal copies of rRNA genes in the nucleolus.

31.2.2. Regulation of Transcription

This mechanism of gene regulation determines whether a given gene will be transcribed or not in a given cell or at a given time. This mode of regulation is universal, so much so that ordinarily the term gene regulation implies regulation of transcription. The various mechanisms of transcription regulation will be discussed in some detail in the subsequent sections.

31.2.3. Post-transcriptional Regulation

Controls after transcription determine if the mRNA produced by a gene is available for translation. For example, a cleavage of primary RNA transcript at a specific site may be required for mRNA function, *e.g.*, the early genes of phage T7. The RNA from early T7 genes is cleaved into 4 monocistronic mRNAs and one dicistronic mRNA. This dicistronic mRNA can be translated only after it is cleaved from the original polycistronic mRNA. The primary transcript of some genes is subject to alternative splicing, *e.g.*, the *tra* gene in *Drosophila*. Whether an embryonic cell will produce Tra protein or not, depends on the pattern of splicing of *tra* gene primary transcript (Section 29.10, Chapter 29).

31.2.4. Translational Regulation

Translational regulation of gene action controls if an mRNA that is suitable for translation will be translated or not. This mode of regulation is based on ribosomes, tRNA, mRNA, regulatory proteins (***autogenous regulation***) and regulatory RNA. For example, a heat labile factor produced in phage-infected cells of *E. coli* combines with the bacterial ribosomes. As a result, these ribosomes preferentially translate phage mRNAs.

In ***autogenous regulation*** a protein or a group of proteins encoded by an mRNA binds to its own mRNA and prevents its translation. For example, ribosomal proteins (r-proteins) become associated, immediately after their production, with rRNA to form ribosomes. If for some reason, r-proteins accumulate in the cell, certain r-proteins bind to their own mRNA and prevent further r-protein synthesis (Fig. 31.2).

31.2.5. Post -Translational Regulation

Post-translational control governs activity of the protein products of genes; it mainly involves protein modification, protein degradation and feed-back inhibition. One or more enzymes of almost all biosynthetic pathways are reversibly inhibited by the end-products of the concerned pathways; this is known as ***feed-back inhibition***. Feed-back inhibition occurs only when the end-product is present above a threshold concentration. When the concentration of the end product declines below this threshold level, feed-back inhibition is automatically reversed or eliminated. The enzymes that show feed-back inhibition are called ***regulatory enzymes***. Usually, regulatory enzymes catalyze the first reaction of the concerned biosynthetic pathways.

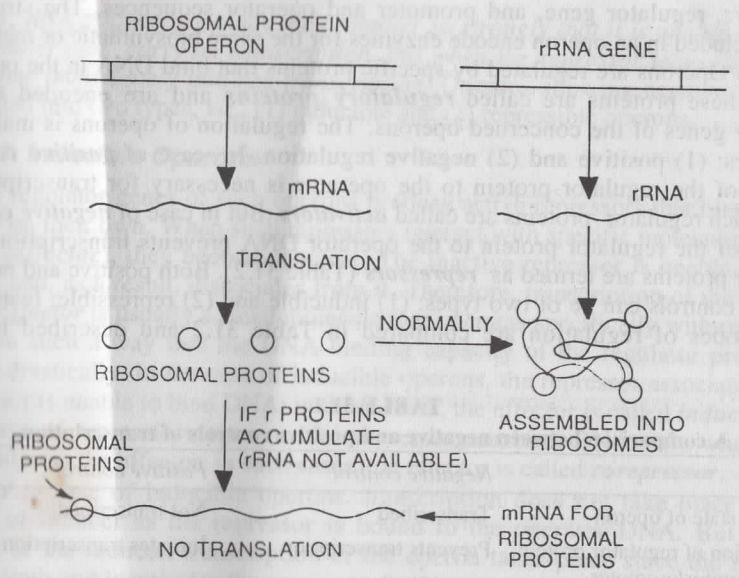


Fig. 31.2. The ribosomal proteins quickly associate with rRNA and together become assembled into ribosomes. But when r-proteins accumulate due to the nonavailability of rRNA for ribosome assembly, the r-proteins bind to the upstream region of their own mRNA; this interaction prevents the translation of the mRNA.

31.3. REGULATION OF TRANSCRIPTION

The evidence leading to the conclusion that gene transcription is regulated was accumulated gradually. It was discovered around 1890 that yeast cells that were cultured in a medium containing lactose began lactose fermentation immediately. In contrast, yeast cells cultured on medium without lactose could begin lactose fermentation only after 14-16 hr.

It was subsequently discovered that *E. coli* cells produced enzymes concerned with lactose metabolism only when lactose was present in the medium. In 1959, Pardee and coworkers concluded that an *E. coli* gene (gene i^+) produces a repressor that prevents the production of enzymes involved in lactose metabolism. It was soon found that the repressor produced by gene i^+ most likely binds to gene o^+ (operator) of *E. coli*. All these findings were synthesized in, now universally accepted operon concept of gene regulation, which is based on gene action regulation by regulation of transcription. The operon concept was proposed by Jacob and Monod in 1961.

Transcription regulation follows similar general schemes in both prokaryotes and eukaryotes, but the details are markedly different. Therefore, a brief account for these groups of organisms is presented separately in Chapters 31 and 32.

31.4. TRANSCRIPTION REGULATION IN PROKARYOTES

In prokaryotes, genes are organized in groups called operons. An ***operon*** consists of a group of structural genes whose transcription is regulated by the same set of

genes, *viz*, regulator gene, and promoter and operator sequences. The structural genes included in an operon encode enzymes for the same biosynthetic or metabolic pathway. Operons are regulated by specific proteins that bind DNA in the operator region; these proteins are called ***regulatory proteins*** and are encoded by the regulator genes of the concerned operons. The regulation of operons is mainly of two types: (1) positive and (2) negative regulation. In case of ***positive control***, binding of the regulator protein to the operator is necessary for transcription to begin; such regulator proteins are called ***activators***. But in case of ***negative control***, binding of the regulator protein to the operator DNA prevents transcription; such regulator proteins are termed as ***repressors*** (Table 31.2). Both positive and negative types of controls can be of two types: (1) inducible and (2) repressible; features of these modes of regulation are compared in Table 31.2 and described in later sections.

TABLE 31.1
A comparison between negative and positive controls of transcription

Feature	*Negative control*	*Positive control*
Default state of operon	Transcribed	Not transcribed
Interaction of regulator protein with operator/promoter	Prevents transcription	Initiates transcription
Inactive regulator protein produces	Recessive constitutivity	Recessive uninducibility or superrepression
Defective operator/DNA binding site produces	Dominant constitutivity	Dominant super-repression/uninducibility

TABLE 31.2
Comparison between inducible and repressible operons

Feature	*Inducible system*	*Repressible system*
Regulator protein (negative state)	Active (binds to operator)	Inactive (does not bind to operator)
Regulator protein after association with effector	Becomes inactive	Becomes active
Transcription of the operon in the presence of effector	Occurs	Does not occur
Effector termed as	Inducer	Corepressor
Active operon is termed as	Induced	Derepressed
An operon that can not become active is called	Uninducible	Super-repressed

In addition, transcription can be regulated by the following mechanisms as well: (1) attenuation, (2) anti-termination and (3) modification of RNA polymerase.

31.4.1. Negative Control

In this system, the regulator protein binds to operator DNA and this binding stops transcription of the concerned operon. Therefore, the regulator protein is called ***repressor***. Transcription of operon can occur only when the repressor dissociates from the operator. The repressor action seems to be affect the binding of RNA polymerase to the promoter region of the operon. *When repressor is bound to the operator, RNA polymerase binds the promoter so strongly that it is unable to leave the promoter and carry on transcription. But when the repressor leaves operator the enzyme immediately initiates transcription.* Operons showing negative control are divided into two types, *viz*., (1) inducible and (2) repressible operons.

31.4.1.1. Inducible Operons

The regulator genes of such operons produce active repressors that bind to the operator on their own. When such repressors interact with specific molecules called ***inducer*** (***effector***), they become inactive. The inactive repressor is unable to bind operator and, as a result, dissociates from it. Therefore, transcription of the operon begins. ***Effector*** is usually a small molecule that interacts specifically with regulator protein in such a way that the DNA binding capacity of the regulator protein is changed drastically. In the case of inducible operons, the repressor associated with the effector is unable to bind DNA; in these cases, the effector is called ***inducer***. But in the case of repressible operons, the repressor can bind DNA only when it is associated with the effector; in such cases, the effector is called ***corepressor***.

Thus in case of inducible operons, transcription does not take place in the absence of inducer as the repressor is bound to the operator DNA. But in the presence of the inducer, transcription of the operon takes place since the inducer interacts with and inactivates the repressor. Such operons are ideal for regulation of production of those enzymes that are involved in substrate utilization. Further, the substrate that is metabolized by the enzymes encoded by such an operon, serves as the inducer for that operon.

The *lac* operon of *E. coli* is an inducible operon. The regulator gene *i* is located at the left extreme of the operon; this is followed by the promoter/operator (*P/O*) region. After the *P/O* region are located three structural genes, *viz*., *z, y* and *a*; these genes encode the enzymes concerned with lactose utilization. The repressor encoded by gene i^+ is active and binds the *O* region; this prevents transcription of *lac* operons. When lactose is present in the medium in which bacterial cells are growing, lactose interacts with the repressor and makes it inactive. The inactive repressor is unable to bind the operator of *lac* operon; therefore, transcription of *lac* operon takes place (Fig. 31.3).

31.4.1.2. Repressible Operons

In case of such operons, the repressor encoded by the regulator gene is inactive, and unable to bind operator. Therefore, the operon is normally functional or ***derepressed***. But when the repressor interacts with the effector, *i.e.*, ***corepressor***, it becomes active and binds operator DNA; as a consequence, transcription of the operon is stopped, *i.e.*, the operon becomes ***repressed***. Such operons generally encode enzymes that catalyze biosynthetic pathways like tryptophan biosynthesis, histidine biosynthesis, etc. Further, the end-product of that biosynthetic pathway, whose enzymes are encoded by an operon, acts as corepressor for that operon.

The *trp* (tryptophan) operon of *E. coli* encodes the enzymes for tryptophan biosynthesis. The regulator gene (*R*) of *trp* operon is located at a distance from the structural genes (Fig. 31.4). The *P/O* sequence is located at the left end of operon, followed by a leader and an attenuator (Section 4.5) sequence, and the five structural genes, *viz*., *trpE, trpD, trpB, trpC* and *trpA*. The repressor encoded by gene *R* is inactive; it can not bind the operator sequence and, as a result, *trp* operon is

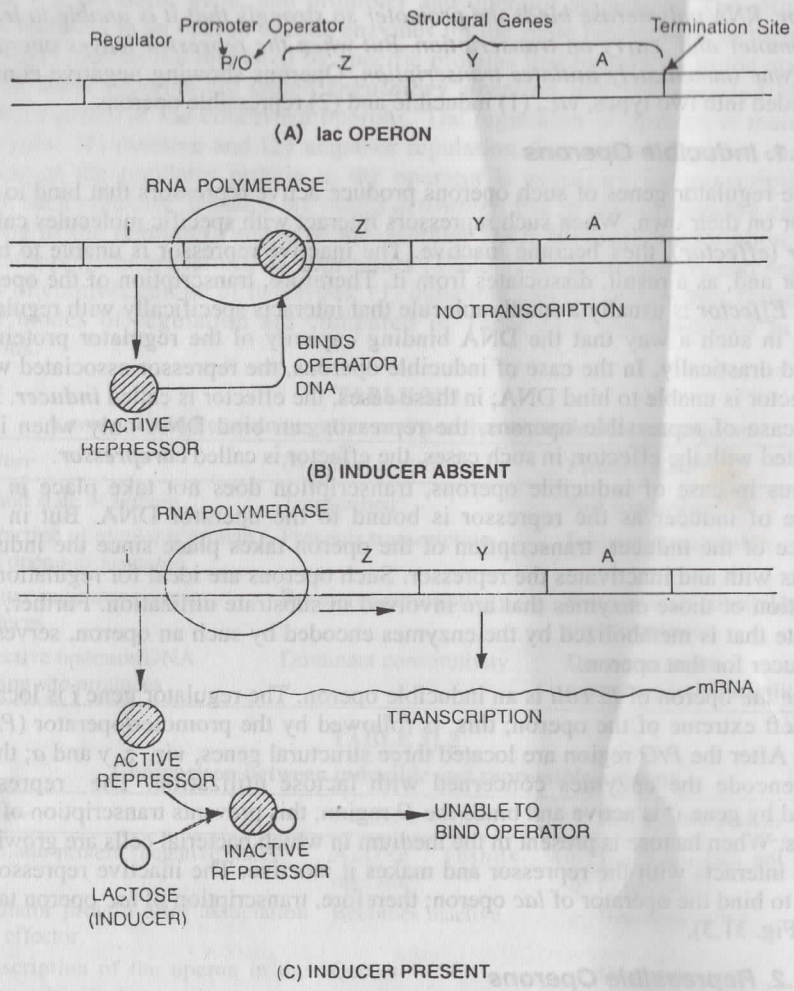


Fig. 31.3. The *lac* operon of *E. coli.* (A) Arrangement of genes in *lac* operon. (B) The repressor produced by regulator gene is in an active form and binds to the operator DNA preventing transcription initiation. (C) But in the presence of inducer, the repressor molecules are inactivated by the inducer molecules so that they can no longer bind to the operator DNA. As a result, the structural genes are transcribed.

normally derepressed, *i.e.*, transcribed. But if, for some reason, tryptophan accumulates in the cell above a threshold level, it interacts with the inactive repressor. As a result, the repressor becomes active and binds the operator of *trp* operon. Therefore, the *trp* operon becomes repressed and transcription is prevented. When the level of tryptophan in the cell declines below a threshold level, the repressor loses its tryptophan, and becomes inactive; as a result, transcription of *trp* operon is resumed (Fig. 31.4).

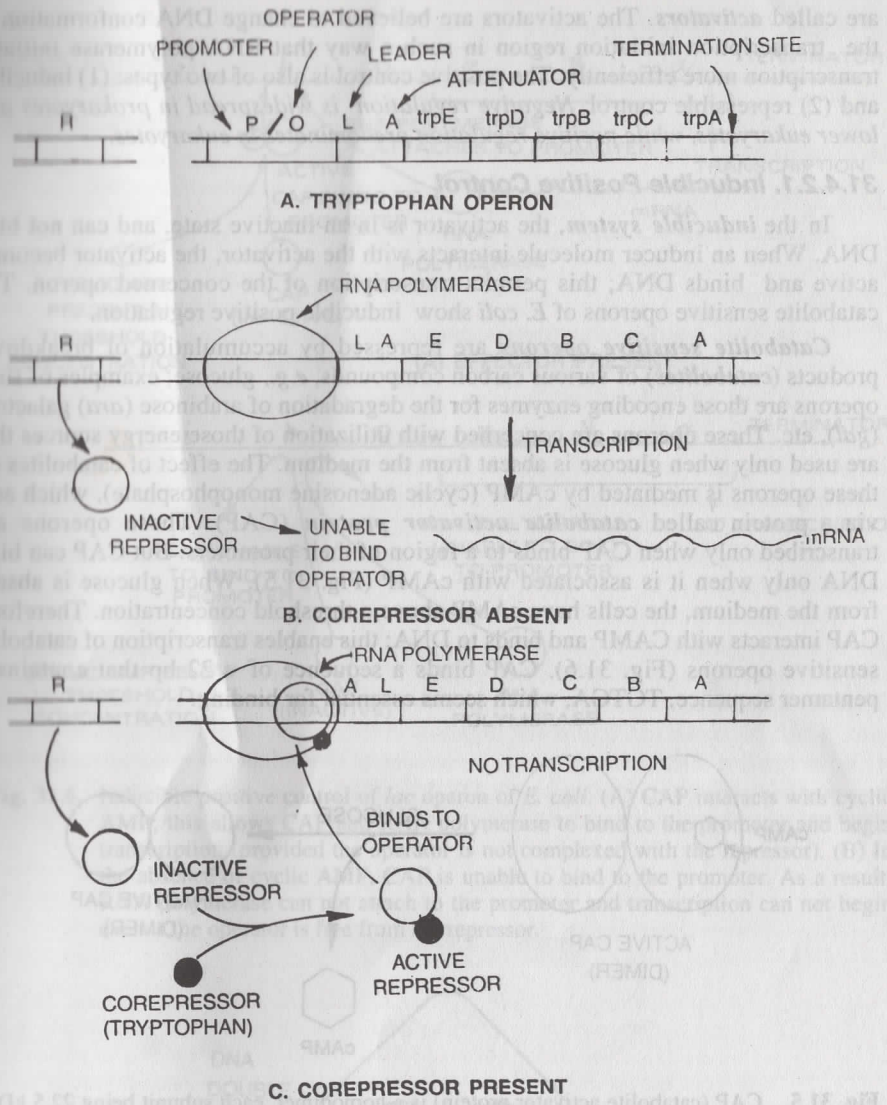


Fig. 31.4. Regulation of the repressible *tryptophan* (*trp*) operon of *E. coli.* (A) Arrangement of the genes in operon. (B) The regulator gene (located at a different site in the chromosome) produces an inactive repressor, which is unable to bind to the operator; hence transcription of the structural genes continues. (C) But when the corepressor *i.e.,* tryptophan, is present at a threshold concentration, it interacts with and activates the repressor, which now binds the operator. This prevents transcription of the operon.

31.4.2. Positive Control

In case of ***positive control***, transcription of the concerned operon does not take place unless the regulator protein is bound to the operator; such regulator proteins

are called ***activators***. The activators are believed to change DNA conformation in the transcription initiation region in such a way that RNA polymerase initiates transcription more efficiently. The positive control is also of two types: (1) inducible and (2) repressible control. *Negative regulation is widespread in prokaryotes and lower eukaryotes, while positive regulation predominates in eukaryotes.*

31.4.2.1. Inducible Positive Control

In the ***inducible system***, the activator is in an inactive state, and can not bind DNA. When an inducer molecule interacts with the activator, the activator becomes active and binds DNA; this permits transcription of the concerned operon. The catabolite sensitive operons of *E. coli* show inducible positive regulation.

Catabolite sensitive operons are repressed by accumulation of breakdown products (***catabolites***) of various carbon compounds, *e.g.,* glucose; examples of such operons are those encoding enzymes for the degradation of arabinose (*ara*) galactose (*gal*), etc. These operons are concerned with utilization of those energy sources that are used only when glucose is absent from the medium. The effect of catabolites on these operons is mediated by cAMP (cyclic adenosine monophosphate), which acts via a protein called ***catabolite activator protein*** (CAP). These operons are transcribed only when CAP binds to a region of their promoters. But CAP can bind DNA only when it is associated with cAMP (Fig. 31.5). When glucose is absent from the medium, the cells have cAMP above a threshold concentration. Therefore, CAP interacts with CAMP and binds to DNA; this enables transcription of catabolite sensitive operons (Fig. 31.6). CAP binds a sequence of ~ 22 bp that contains a pentamer sequence, TGTGA, which seems essential for binding.

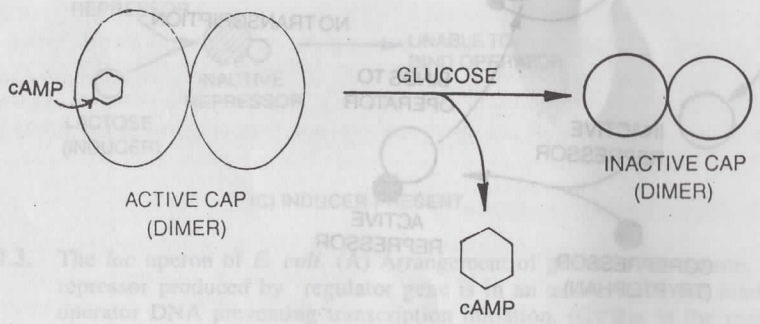


Fig. 31.5. CAP (catabolite activator protein) is a homodimer, each subunit being 22.5 kDA. Association of the dimer with a single cAMP molecule makes it active. Presence of glucose in the medium causes cAMP depletion and CAP inactivation.

But in the presence of glucose in the medium, glucose is preferentially used by cells. This is because glucose somehow reduces cAMP level in the cell below the threshold level needed for CAP activation. As a result, CAP is unable to bind the promoters of 'catabolite sensitive' operons and they are not transcribed. The promoters of these operons are defective in having poorly conserved –35 sequence and, in some cases, even –10 sequence. CAP - binding induces a > 90° bend in DNA molecules (Fig. 31.7); this somehow enhances transcription initiation.

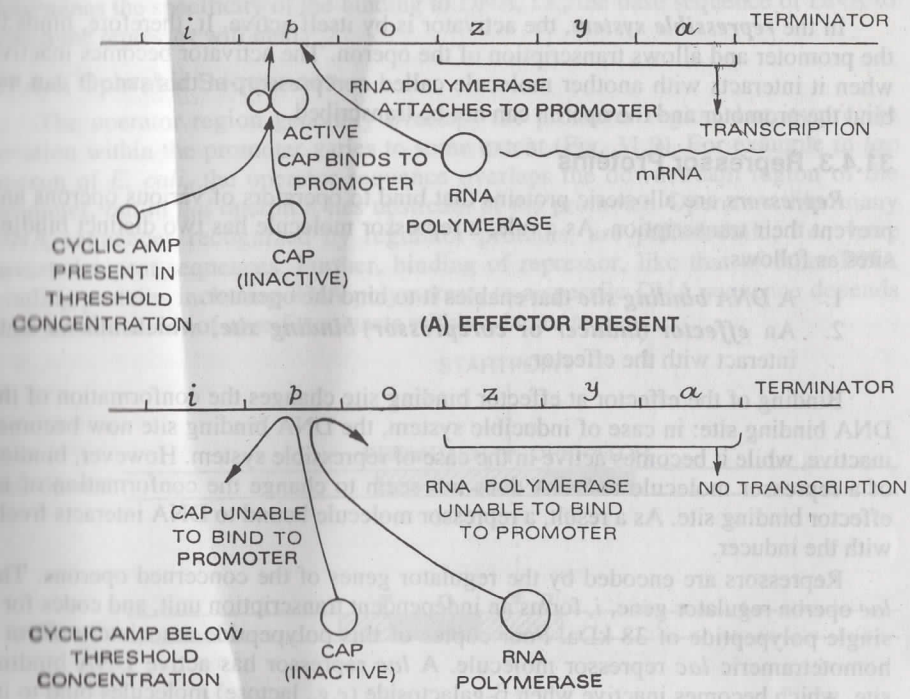


Fig. 31.6. Inducible positive control of *lac* operon of *E. coli*. (A) CAP interacts with cyclic AMP; this allows CAP and RNA polymerase to bind to the promoter and begin transcription (provided the operator is not complexed with the repressor). (B) In the absence of cyclic AMP, CAP is unable to bind to the promoter. As a result, RNA polymerase can not attach to the promoter and transcription can not begin even if the operator is free from the repressor.

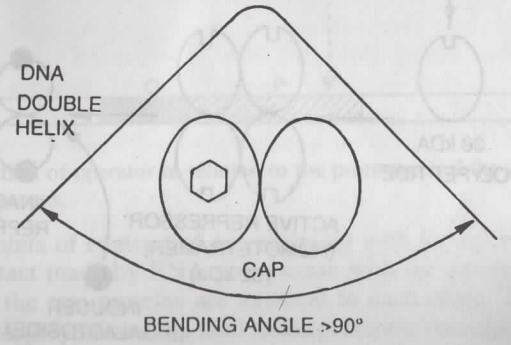


Fig. 31.7. Biding of CAP bends the DNA >90° around the centre of symmetry of binding site

31.4.2.2. Repressible Positve Control

In the ***repressible system***, the activator is by itself active. It, therefore, binds to the promoter and allows transcription of the operon. The activator becomes inactive when it interacts with another molecule called corepressor; in this state it can not bind the promoter and the operon can not be transcribed.

31.4.3. Repressor Proteins

Repressors are allosteric proteins that bind to operators of various operons and prevent their transcription. As a rule, a repressor molecule has two distinct binding sites as follows.

1. A ***DNA binding site*** that enables it to bind the operator.
2. An ***effector (inducer*** or ***corepressor) binding site***, which allows it to interact with the effector.

Binding of the effector at effector binding site changes the conformation of the DNA binding site: in case of inducible system, the DNA binding site now becomes inactive, while it becomes active in the case of repressible system. However, binding of a repressor molecule to DNA does not seem to change the conformation of its effector binding site. As a result, a repressor molecule bound to DNA interacts freely with the inducer.

Repressors are encoded by the regulator genes of the concerned operons. The *lac* operon regulator gene, *i*, forms an independent transcription unit, and codes for a single polypeptide of 38 kDa. Four copies of this polypeptide associate to form a homotetrameric *lac* repressor molecule. A *lac* repressor has active DNA binding site, which becomes inactive when β-galactoside (*e.g.*, lactose) molecules bind to its effector binding site. Binding of two molecules of the inducer to one molecule of the repressor is enough to inactivate the DNA binding site of the repressor (Fig. 31.8). The DNA binding site of *lac* repressor forms a short α-helix; this feature is common

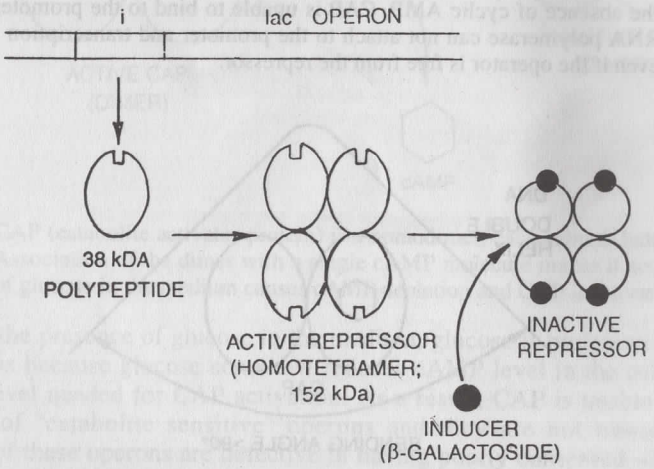


Fig. 31.8. The *lac* repressor is a homotetramer. It undergoes conformational change when inducer molecules bind to it: binding of 2 molecules of inducer to one repressor (tetramer) is enough to release repression.

to many other bacterial regulator proteins. The amino acid sequence of this α–helix determines the specificity of the binding to DNA, *i.e.*, the base sequence of DNA to which the repressor will bind.

31.4.4. Operator Sequences

The operator region generally overlaps the promoter region, but its precise location within the promoter varies to some extent (Fig. 31.9). For example in *lac* operon of *E. coli*, the operator sequence overlaps the downstream region of the promoter, but in *gal* operon it lies upstream of the promoter. Operators, like many DNA sequences recognized by regulator proteins, are palindromic, *i.e.*, have inverted repeat sequences. Further, binding of repressor, like that of other DNA binding proteins including RNA polymerase, to a specific DNA sequence depends on a small number of specific contacts within a larger region.

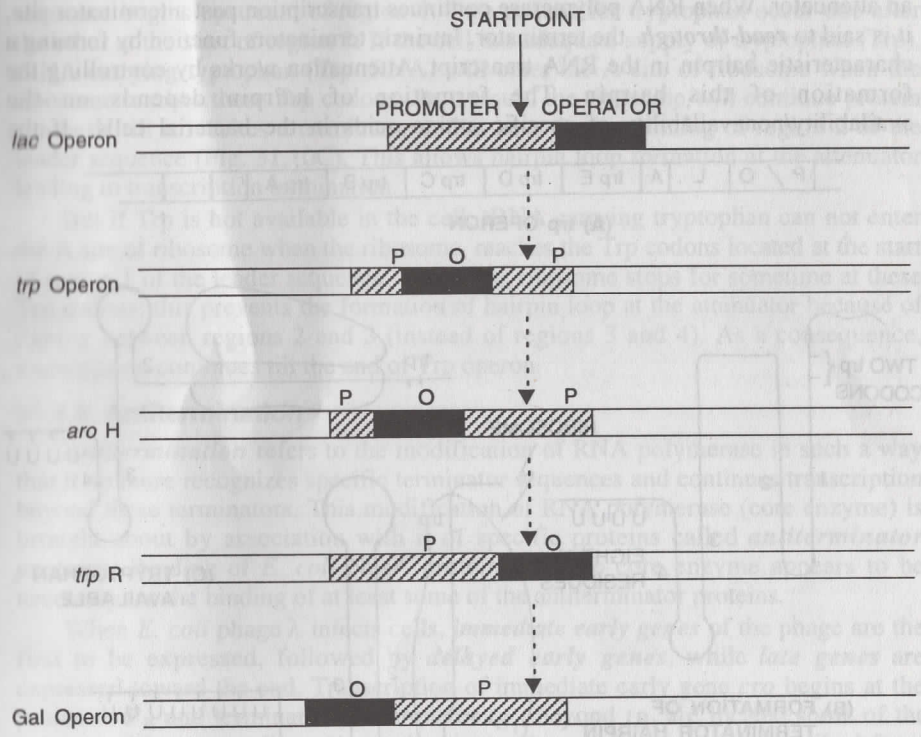


Fig. 31.9. The location of operator in relation to the promoter and the startpoint in different *E. coli* operons.

The major points of contract of *lac* repressor with *lac* operator do not overlap the points of contact made by RNA polymerase with *lac* promoter; the regions of contact made by the two proteins are adjacent to each other. Therefore, both *lac* repressor and RNA polymerase can bind to the promoter/operator region of the same *lac* operon at the same time. In fact, binding of the repressor to *lac* operator seems to increase the strength of binding of RNA polymerase to the adjacent *lac* promoter sequence. Therefore, binding of repressor to *lac* operator stores RNA polymerase at

the *lac* promoter; as a result, transcription starts immediately after the repressor is removed from the operator due to an interaction with the inducer.

In the presence of repressor at the *lac* operator, RNA polymerase binds to the promoter and remains as closed binary complex. But once the repressor is removed, the closed compelx is convested into an open binary complex three-times more rapidly than in the total absence of the repressor. Thus *lac* repressor appears to ensure a far more speedy initiation of transcription once lactose becomes available to the *E. coli* cells.

31.4.5. Attenuation

Transcription units subject to attenuation contain an attenuator. An ***attenuator*** is an intrinsic terminator located at the begining of a transcription unit. The phenomenon of ***attenuation*** controls the ability of RNA polymerase to read-through an attenuator. When RNA polymerase continues transcription past a terminator site, it is said to ***read-through*** the terminator. Intrinsic terminators function by forming a characteristic hairpin in the RNA transcript. Attenuation works by controlling the formation of this hairpin. The formation of hairpin depends on the availability/nonavailability of specific amino acids in the bacterial cells. If the

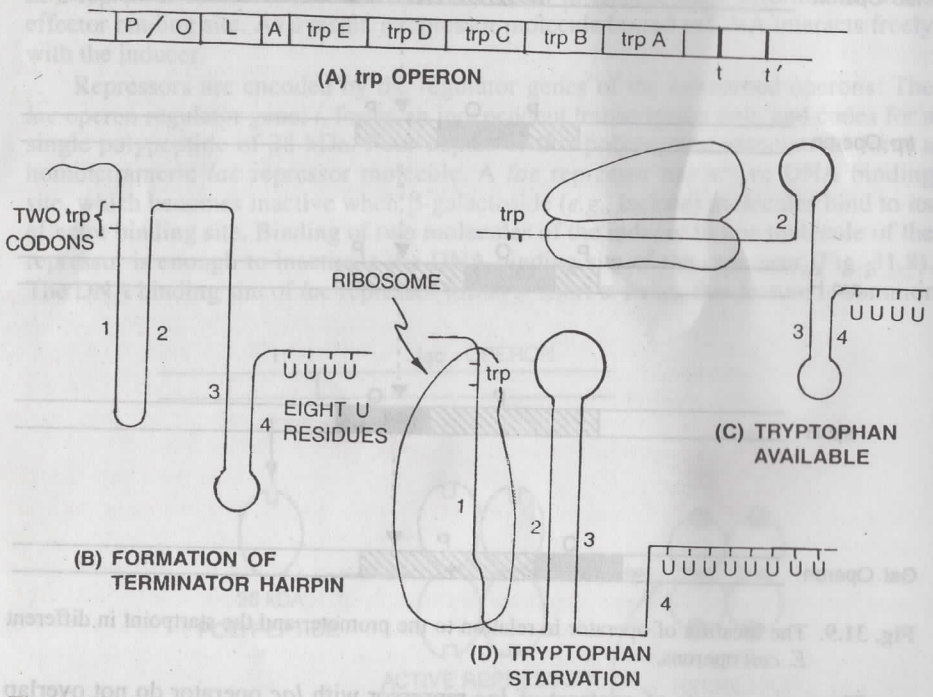


Fig. 31.10. The mechanism of attenuation of the *trp* operon of *E. coli.* (A) The first structural gene of the operon, *trpE,* is preceded by the *P* (promoter), *O* (operator), *L* (leader) and *A* (attenuator) sequences. (B) The base sequence allows the leader to form a secondary structure, which has terminator hairpin. (C) The leader sequence yields a typical termination hairpin signal with the run of U residues if tryptophan is available in the cell. (D) But when Trp is absent from the cell, the termination does not occur since the hairpin does not form.

concerned amino acid is not available, the hairpin does not from in the RNA transcript at the attenuator, and as a result, transcription continues. However, if the concerned amino acid is available in the cell, the hairpin forms at the attenuator. As a consequence, transcription stops before it reaches the first structural gene of the operon. Examples of operons subjected to attenuation are *trp, thr, his* and some other operons of *E. coli*. In case of *his* operon, attenuation is the only mechanism of transcription regulation.

The *trp* operon of *E. coli* has a leader (*L*) and an attenuator (*A*) sequence, in that order, immediately downstream of the operator sequence (Fig. 31.10A). RNA transcript of the leader has 4 regions, designated as 1, 2, 3 and 4. Ordinarily region 1 base-pairs with region 2, and region 3 pairs with region 4 (Fig. 31.10B). Pairing between regions 3 and 4, generators a hairpin loop followed by 8 U residues in the attenuator region, which signals transcription termination. Translation of leader sequence begins upstream of region 1. Two codons for tryptophan occur one after the other at the start of region 1. If the cell has adequate supply of tryplophan (Trp), tRNA carrying typtophan (Trp-tRNA) will enter the A site of ribosome when the ribosome reaches these Trp codons. As a result, the ribosome will continue protein synthesis till it reaches a terminator codon near the benginning of region 2 of the leader sequence (Fig. 31.10C). This allows hairpin loop formation at the attenuator leading to transcription termination.

But if Trp is not available in the cell, tRNA carrying tryptophan can not enter the A site of ribosome when the ribosome reaches the Trp codons located at the start of region 1 of the leader sequence. Therefore, ribosome stops for sometime at these Trp codons; this prevents the formation of hairpin loop at the attenuator because of pairing between regions 2 and 3 (instead of regions 3 and 4). As a consequence, transcription continues till the end of Trp operon.

31.4.6. Antitermination

Antitermination refers to the modification of RNA polymerase in such a way that it no more recognizes specific terminator sequences and continues transcription beyond these terminators. This modification of RNA polymerase (core enzyme) is brought about by association with it of specific proteins called ***antiterminator proteins***. Binding of *E. coli* NusA protein with the core enzyme appears to be necessary for the binding of at least some of the antiterminator proteins.

When *E. coli* phage λ infects cells, ***immediate early genes*** of the phage are the first to be expressed, followed by ***delayed early genes***, while ***late genes*** are expressed toward the end. Transcription of immediate early gene *cro* begins at the promoter P_R and terminates at terminator t_{R_1}; beyond t_{R_1} are located some of the delayed early genes. Gene *N*, another immediate early gene, is transcribed from another promoter P_L, and encodes protein *p*N, which functions as antiterminator protein. So long as *p*N is not produced, transcription of *cro* terminates at t_{R_1}.

Gene *cro* contains a sequence called *nut* site, which is a 17 base conserved sequence. When *p*N has been produced by λ immediate eary gene *N, p*N attaches to the core enzyme as it is transcribing the *nut* region of gene *cro.* Protein *p*N possibly recognizes the *nut* sequence of RNA transcript being produced, and then binds the RNA polymerase engaged in transcription. *p*N interacts with the β subunit of core enzyme and prevents termination at t_{R_1} (Fig. 31.11). Protein *pN* can bind core enzyme only when *E. coli* protein NusA is bound to the core enzyme.

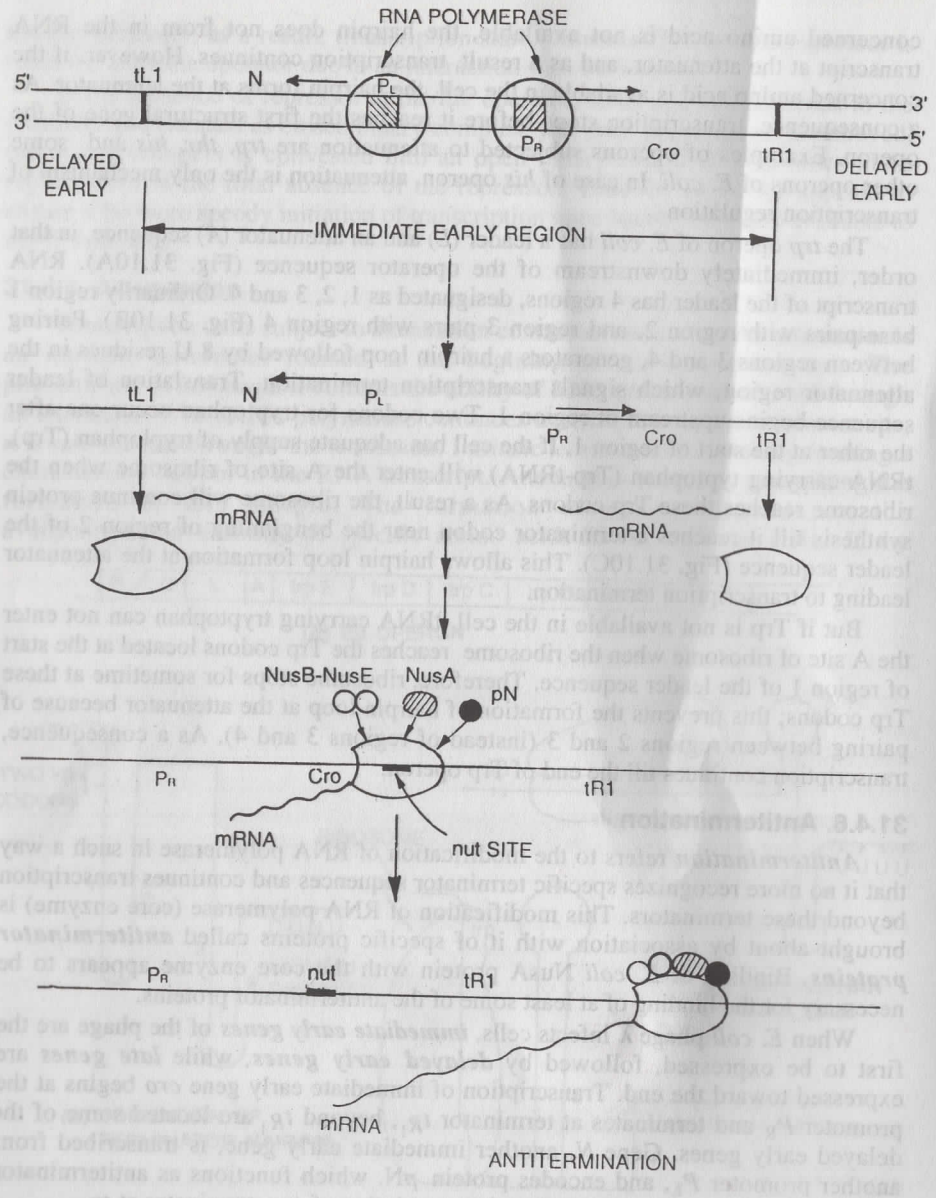


Fig. 31.11. Antitermination by pN (●) in phage λ. pN, along with the bacterial proteins NusA (⊘) and NusB-NusE dimer (OO), assocates with the core enzyme as it passes the *nut* site located within the transcripton unit.

31.4.7. Alternative Forms of Bacterial RNA polymerase

The bacterial RNA polymerase seems to alternate between an initiation competent and a termination competent from (Fig. 31.12).

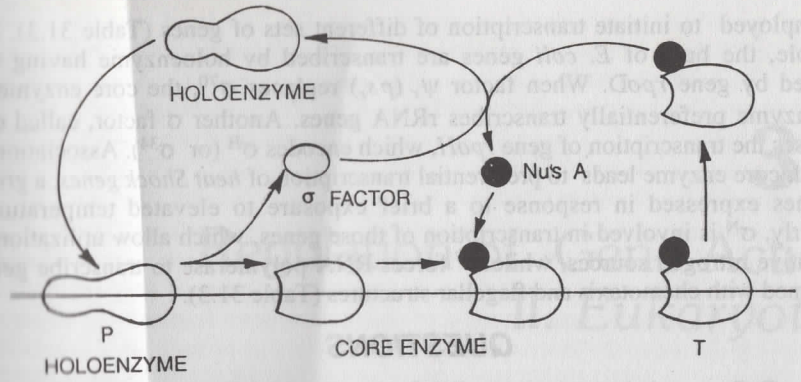


Fig. 31.12. The two alternative forms of bacterial RNA polymerase. P, promoter; T, terminator. The movement of RNA polymerase through the transcription unit is shown in a single figure.

31.4.7.1. Initiation Competent Form

This form of the enzyme is represented by the holoenzyme (α_2 β β'σ), which specifically binds promoters. This form initiates transcription but as the holoenzyme leaves the promoter region, σ factor dissociates and leaves the core enzyme. The core enzyme carries on with the transcription.

31.4.7.2. Termination Competent Form

This form is produced by association of *E. coli* protein NusA with the core enzyme after σ factor has dissociated from the holoenzyme, following initiation of transcription. The α_2ββ′.NusA form of enzyme has a tendency to pause during transcription; this enhances the efficiency of termination at terminator sites. Protein NusA remains associated with the core enzyme as long as it is associated with DNA. But when termination occurs, σ factor displaces NusA from the core enzyme, and again produces the holoenzyme (Fig. 31.12).

31.4.8. Modification of RNA Polymerase

In some cases, transcription regulation may rely on a modification of RNA polymerase. The modification is usually based on σ factor, but in few cases it may involve a modification of the core enzyme itself. The ***sigma factor*** often serves as an effective mechanism to selectively transcribe different genes, *i.e.*, different σ factors

TABLE 31.3

Some σ factors of *E. coli* and the genes preferentially transcribed under their influence

Sigma factor	*Genes preferentially transcribed*
σ^{70}	Bulk of genes.
σ^{24}	*rpo*^{H} (encodes σ^{H}).
σ^{F}	Genes for chemotaxis and flagellar structures.
σ^{H}	Heat shock genes
σ^{N}	Genes for utilization of alternative nitrogen sources.
ψ_r (*psi*$_r$)	*r*RNA genes.

are employed to initiate transcription of different sets of genes (Table 31.3). For example, the bulk of *E. coli* genes are transcribed by holoenzyme having σ^{70} encoded by gene *rpo*D. When factor ψ_r (ps_r) replaces σ^{70}, the core enzyme-ψ_r holoenzyme preferentially transcribes rRNA genes. Another σ factor, called σ^{24}, increases the transcription of gene *rpoH*, which encodes σ^H (or σ^{34}). Association of σ^H with core enzyme leads to preferential transcription of *heat Shock genes*, a group of genes expressed in response to a brief exposure to elevated temperatures. Similarly, σ^N is involved in transcription of those genes, which allow utilization of alternative nitrogen sources, while σ^F forces RNA polymerase to transcribe genes concerned with chemotaxis and flagellar structures (Table 31.3).

QUESTIONS

1. Define regulation gene action. Briefly describe the different levels at which gene action is regulated in the biological world.
2. Define operon. Explain the operon model of gene regulation using *lac* operon of *E. coli* as an example.
3. Explain the mechanisms of inducible and repressible operons with the help of suitable examples.
4. List the various strategies for regulation of transcription in prokaryotes, and explain the strategies other than the operon system.
5. Write short note on the following :
   (*i*) operon, (*ii*) repressor,
   (*iii*) effector, (*iv*) operator,
   (*v*) attenuation, (*vi*) antitermination,
   (*vii*) positive regulation, (*viii*) negative regulation,
   (*x*) σ-factor.

32

Regulation of Gene Action II. *Eukaryotes*

32.1. INTRODUCTION

The mechanism of gene regulation in eukaryotes is not as well understood as that in prokaryotes, but the general mode of gene regulation appears to be similar in both. Both regulator and promoter genes are known in eukaryotes, but eukaryotes do not have typical operons like prokaryotes. In eukaryotes, gene action may be regulated at the level of (1) activation of gene structure, (2) transcription, (3) translation, (4) gene replication (amplification and destruction, (5) after transcription (RNA processing and transport into the cytoplasm), and (6) after translation. The regulations of translation and transcription are considered in some detail in this chapter. But before this, it would be rewarding to consider some well studied phenomenon concerning gene expression in eukaryotes; these phenomena are: (1) position effects, (2) controlling elements, (3) paramutation, (4) dosage compensation, (5) genetic imprinting and (6) epigenetic inheritance.

32.2. POSITION EFFECTS

Position effect represents a change in the function of a gene due to its position in a chromosome or in the homologous chromosome. Thus a position effect does not represent a change in the primary structure (base sequence) of a gene. Position effects are known in yeast, *Neurospora,* maize, *Oenothera, Drosophila,* mice, man, etc. The various cases of position effects may be classified into the following three categories: (1) variegated position effects, (2) *cis-trans* position effects and (3) duplication position effects.

32.2.1. Variegated Position Effects

Variegated position effects involve heterochromatinization, which prevents the transcription of concerned genes. Heterochromatin is not only less active in transcription, it has more repetitive DNA sequences than euchromatin. Heterochromatin is normally located in specific positions, generally near centromeres and telomeres, of different chromosomes; such heterochromatin is termed as ***constitutive heterochromatin.*** When an euchromatic segment is placed close to heterochromatin, it undergoes heterochromatinization, *i.e.,* it changes into heterochromatin; such a heterochromatin is known as ***facultative heterochromatin.***

An example of variegated position effect due to heterochromatinization is provided by the mottled eye mutant in *Drosophila*. The gene w^+ is normally located in 3C2 band of the *X* chromosome, near the tip of the left arm, and produces dull red

eye colour. When a paracentric inversion places the 3*C*2 band (and the w^+ allele) near or in the heterochromatic region close to the centromere, a mottled eye colour results. The mottled eye mutant does not involve a change in the base sequence of the w^+ gene. This is evident from the fact that a dull red eye is produced if another inversion places the 3*C*2 band in its original position. Further, when a normal w^+ allele is placed in the position of the mottled allele by means of a crossing over, the new w^+ allele also produces a mottled eye (not the expected dull red eye).

It appears that when an euchromatic segment is brought close to a heterochromatic region of the chromosome, it undergoes heterochromatinization, and there is a consequent reduction in template activity of this segment. The frequency of heterochromatinization of the euchromatic segment depends on its location in relation to the heterochromatin (the closer it is to the heterochromatin, the more frequent is heterochromatinization) as well as on its size (the smaller the euchromatic segment, the greater the frequency of heterochromatinization). Conversely, when a heterochromatic segment is inserted into an euchromatic segment, it undergoes euchromatization.

Different heterochromatic regions may differ in their ability to induce heterochromatinization. For example, band 3*C*2 shows a relatively lower frequency of heterochromatinization when placed near the heterochromatin in the *X* chromosome than when placed near that in the chromosome IV. Heterochromatinization of a chromosome segment may be affected by the sex of the parent, which contributed this chromosome. In female *Drosophila,* paternal *X* chromosome appears to be heterochromatinized in 71 per cent of nuclei, while the maternal *X* chromosome is heterochromatinized in only 20% of nuclei.

Heterochromatinization may also be affected by other chromosomes in the cell. For example, the presence of an additional *Y* chromosome in *Drosophila* reduces the frequency of heterochromatinization and, as a result, the variegation of a trait. Thus heterochromatinization and euchromatinization are reversible changes; they most likely represent differences in the transcription activity due to a greater coiling of chromatin fibers in the heterochromatic than in the euchromatic state.

The overall properties of chromatin are influenced by a group of ATPases, called ***SMC (structural maintenance of chromatin) proteins,*** which are divided into the following two functional groups: (1) condensins and (2) cohesins. ***Condensins*** are involved in the control of overall chromatin structure. The heterodimer SMC2·SMC4 forms a core that associates with other nonSMC proteins of the condensin complex. ***Cohesins*** are concerned with connections between sister chromatids that must be released during mitosis. Cohesins have the core of SMC1 · SMC3 heterodimer, which associates with other nonSMC proteins to yield the cohesin complex. The condensin complex is specifically activated at mitosis; it introduces positive supercoils into DNA by using ATP hydrolysis and topisomerase I, and causes chromatin condensation *in vitro*.

The condensation state of interphase heterochromatin is close to that of mitotic chromosomes. The heterochromatin is transcriptionally repressed, replicates late in S phase and may be localized to the nuclear periphery. Formation of heterochromatin is not completely dependent on DNA sequence so that when a gene is placed adjacent to heterochromatin it may become inactive. The inactivation spreads from heterochromatin into the adjacent region to a variable distance. Therefore, the closer

a gene lies to heterochromatin, the greater is the probability of its inactivation. Thus formation of heterochromatin occurs in two steps: ***nucleation***, *i.e.,* the initiation of heterochromatin formation occurs at a specific DNA sequence, while ***propagation***, *i.e.,* spread of heterochromatin organization, continues to variable distances possibly depending upon the quantities of limiting protein components. Thus the boundaries of heterochromatic regions are defined by exhaustion of one of the proteins required for heterochromatin formation. This phenomenon explains position effect variegation, *e.g.,* in *Drosophila*, and telomeric silencing observed in yeast.

Mating type locus in yeast also shows gene inactivation most likely by heterochromatin formation; the gene products involved in this inactivation also participate in telomeric silencing. According to one model, one of these proteins, RAP1, is a sequence-specific DNA-binding protein; it binds to $C_{1-3}A$ repeats at telomeres, and to the *cis*-acting silencer elements needed for repression of *HML* and *HMR* at the mating type locus. Proteins SIR3 and SIR4 possibly form a heterodimer SIR3 · SIR4, which interacts with RAP1, and with the N-terminal tails of histones H3 and H4. Thus RAP1 recognizes the sequence at which heterochromatin is to form, and recruits SIR3/SIR4 to the site. Once SIR3 · SIR4 has bound to histones H3/H4, the complex may polymerize further and spread along the chromatin fiber. The C-terminus of SIR3 has a similarity to nuclear lamin proteins, and may be responsible for tethering heterochromatin to the nuclear periphery.

32.2.2. *Cis-trans* Position Effects

The ***cis-trans position effect*** is of relatively common occurrence. For example, *X*-linked loci *apr* (*apricot*) and *w* (*white*), governing eye colour in *Drosophila* are closely linked and produce dull red eyes in the heterozygous females with the genetic constitution apr^+ w^+/apr *w, i.e.,* the wild type alleles of the two genes are present in the same chromosome (**cis-*position***). However, when the heterozygous females are apr^+ *w/apr* w^+, *i.e.,* the wild type alleles of the two genes are present in the homologous chromosomes (**trans-*position***), the resulting eye colour is pale apricot. Thus the phenotypic effect of *apr* and *w* genes in the heterozygous state depends on their *cis-* or *trans*-position, hence it is known as **cis-trans *position effect.***

Such position effects may be explained on the basis of the following hypothesis. The polypeptides produced by the two genes showing *cis-trans* position effect, *e.g., apr* and *w,* combine together to form the functional protein. If the two genes are transcribed into a single mRNA molecule, or if translation occurs at the site of transcription, the polypeptides produced by the two genes present in the same chromosome will preferentially combine with each other due to their closeness in space. As a result, an apr^+ w^+ chromosome will produce a functional protein since both the polypeptides produced by the chromosome will be normal. But the chromosomes apr^+ *w* and *apr* w^+ will produce nonfunctional proteins since these protein molecules would have a defective polypeptide produced by either the *apr* or the *w* alleles.

32.2.3. Duplication of Chromosome Segment

Position effect produced by duplication resulting from unequal crossing over is clearly demonstrated in the case of *Bar* (*B*) eye gene of *Drosophila*. However, the molecular mechanism of position effects in such cases is not known. Position effects

would generally be classified as mutational changes unless cytogenetic and genetic studies would demonstrate otherwise.

32.3. PARAMUTATION

Paramutation refers to a change in the function of an allele of a structural gene (***paramutable allele***) that may last for several generations, and is induced by its presence in the same nucleus with another allele, the ***paramutagenic allele,*** of the gene. For example, gene R^r in chromosome 10 of maize produces full anthocyanin in aleurone and certain vegetative parts. When R^r is maintained in the heterozygous state with its allele R^{st}, the R^r allele recovered from the heterozygote has a reduced capacity for anthocyanin production. The reduced function of R^r persists for several generations, and shows a partial recovery when it is maintained in the homozygous state, in the heterozygous state with its *r* allele, or in a hemizygous state (the homologous chromosome 10 being deficient for the R^r locus). *Thus the paramutated state of an allele is induced by a paramutagentic allele, while the recovery in the paramutated allele is spontaneous.*

Paramutation has been extensively studied at *b, r,* and *pl* loci of maize; these genes regulate the synthesis of red and purple pigments by regulating transcription of the anthocyanin biosynthetic genes. Paramutation is produced due to an interaction between two specific alleles of a paramutable gene; the resulting epigenetic change in the allele occurs at a very high frequency and is transmissible through meiosis. An ***epigenetic change*** is an alteration in gene expression associated with alternative chromatin or methylation states (or both) *without any change in DNA sequence*. The alternative epigenetic states are frequently heritable through mitosis and, sometimes, through meiosis, *e.g.,* paramutation; for some reason, transmission through meiosis is more common in plants than it is in animals. Several mechanisms have been proposed to explain paramutation, the following being the important ones.

1. The paramutagenic allele causes an increase in the number of copies of a heterochromatic sequence, called ***metamere,*** that is associated with the paramutable allele.
2. The sequence of paramutable allele changes, perhaps in a subtle but critical region, by a process of gene conversion.
3. The paramutable allele undergoes methylation (at C residues) under the influence of paramutagenic allele.
4. The chromatin structure at paramutable allele may change due to an interaction with the paramutagenic allele. For example, pro' ,ns bound to a regulatory sequence on one allele might interact with those bound to the homologous allele.

The current view of paramutation tends to favour C methylation and changed chromatin structure models. Recently, a maize gene (*Mop1, mediator of paramutation 1*) has been identified that is required for establishment and maintenance of the paramutagenic state of several loci.

32.4. GENETIC IMPRINTING

There is considerable evidence that different sets of genes may be in an inactive state in the male and female gametes of several organisms, including man, mouse, etc.

This leads to the phenomenon called ***genetic imprinting,*** or simply ***imprinting,*** which describes a difference in the behaviour or activity between the alleles contributed by the two parents of an individual. It may be expected that the set of genes involved in genetic imprinting would be such that the maternal and paternal chromosomes present in a zygote complement each other so as to ensure normal differentiation and development. Genetic imprinting is postulated to be due to differential states of methylation, *e.g.,* of CpG 'doublets', of the concerned genes in the male and female gametes.

An example of genetic imprinting is provided by the gene encoding IGF-II (insulin-like growth factor II) in mice. The expression of the alleles of IGF-II gene (and several other genes) in embryos depends upon the sex of the parent from which they were inherited. The IGF-II allele inherited from the father is expressed, while that contributed by the mother is not. It is explained on the basis of methylation of

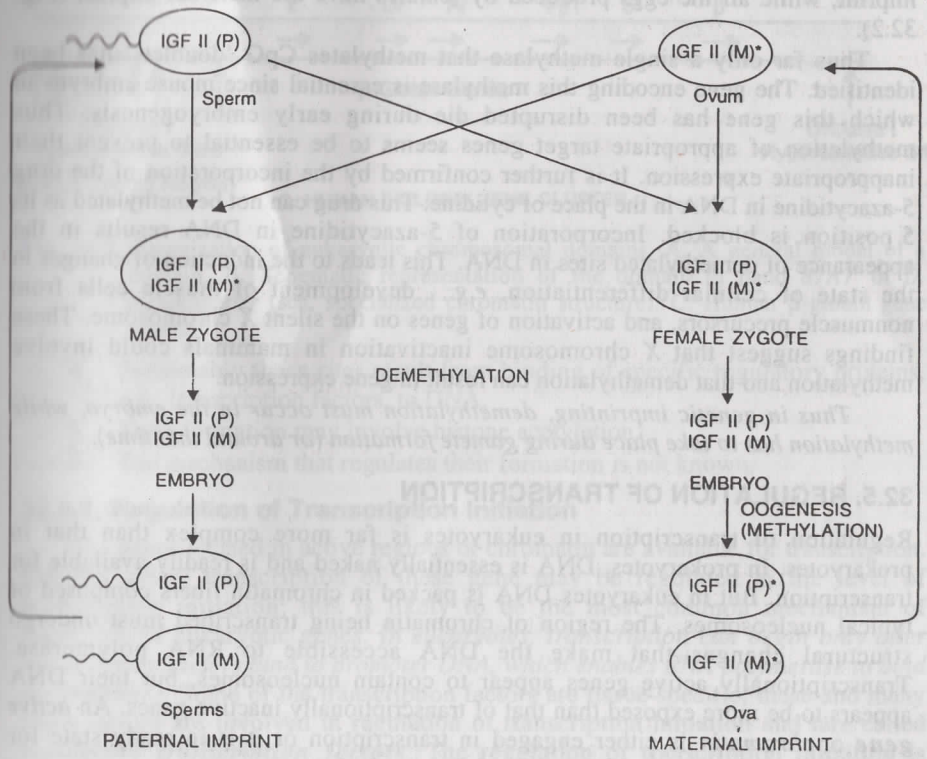


Fig. 32.1. Genetic imprinting in mice. IGF II allele contributed by male gamete [denoted as IGF II(P)] is nonmethylated and expressed, while that by female is methylated [represented by IGF II(M)*] and is not expressed.

the IGF-II allele present in eggs, and nonmethylation of that present in the sperms; as a result, the allele contributed by the sperm is expressed in the zygote (Fig. 32.2). This is the most common pattern in genetic imprinting, but in some cases, the allele contributed by the female is expressed.

This explanation requires that the pattern of methylation is established during gametogenesis and this pattern is reversible. The methylation state of individual genes seems to be reset during gametogenesis and early development. The previous patterns of methylation are erased in the primordial germ cells as they develop in the embryo so that the previous genetic imprints present in the two alleles in these cells are erased, and a typical gene is then in nonmethylated state. In case of females, the maternal pattern of methylation is established during oogenesis on both the alleles of a female. But in case of males, the methylation patterns develops in two stages : (1) prior to spermatogenesis (sperms have the same methylation pattern as spermatocytes) and (2) after fertilization (the methylation pattern is further modified). Therefore, all the sperms produced by the males carry the paternal imprint, while all the eggs produced by females have the maternal imprint (Fig. 32.2).

Thus far only a single methylase that methylates CpG 'doublets' has been identified. The gene encoding this methylase is essential since mouse embryos in which this gene has been disrupted die during early embryogenesis. Thus methylation of appropriate target genes seems to be essential to prevent their inappropriate expression. It is further confirmed by the incorporation of the drug 5-azacytidine in DNA in the place of cytidine. This drug can not be methylated as its 5 position is blocked. Incorporation of 5-azacytidine in DNA results in the appearance of demethylated sites in DNA. This leads to the induction of changes in the state of cellular differentiation, *e.g.,* development of muscle cells from nonmuscle precursors, and activation of genes on the silent *X* chromosome. These findings suggest that *X* chromosome inactivation in mammals could involve methylation and that demethylation can result in gene expression.

Thus in genetic imprinting, demethylation must occur in the embryo, while methylation has to take place during gamete formation (or around that time).

32.5. REGULATION OF TRANSCRIPTION

Regulation of transcription in eukaryotes is far more complex than that in prokaryotes. In prokaryotes, DNA is essentially naked and is readily available for transcription. But in eukaryotes DNA is packed in chromatin fibers composed of typical nucleosomes. The region of chromatin being transcribed must undergo structural changes that make the DNA accessible to RNA polymerase. Transcriptionally active genes appear to contain nucleosomes, but their DNA appears to be more exposed than that of transcriptionally inactive genes. An ***active gene*** or ***chromatin*** is either engaged in transcription or it is in ready state for transcription. Active chromatin has the following features.

1. It has ~200 bp regions or sites at which DNA is 100 times more sensitive to DNAase I (and other nucleases, and to chemical agents) than is the bulk of chromatin; these sites are called ***DNAase I hypersensitive sites***. The region of hypersensitivity extends several kilobases.
2. DNAase I hypersensitive sites lack typical nucleosome organization.

3. Such sites appear in the 5′ regions of all genes before their transcription begins; most likely they are a prerequisite for initiation of transcription (Fig. 32.2)

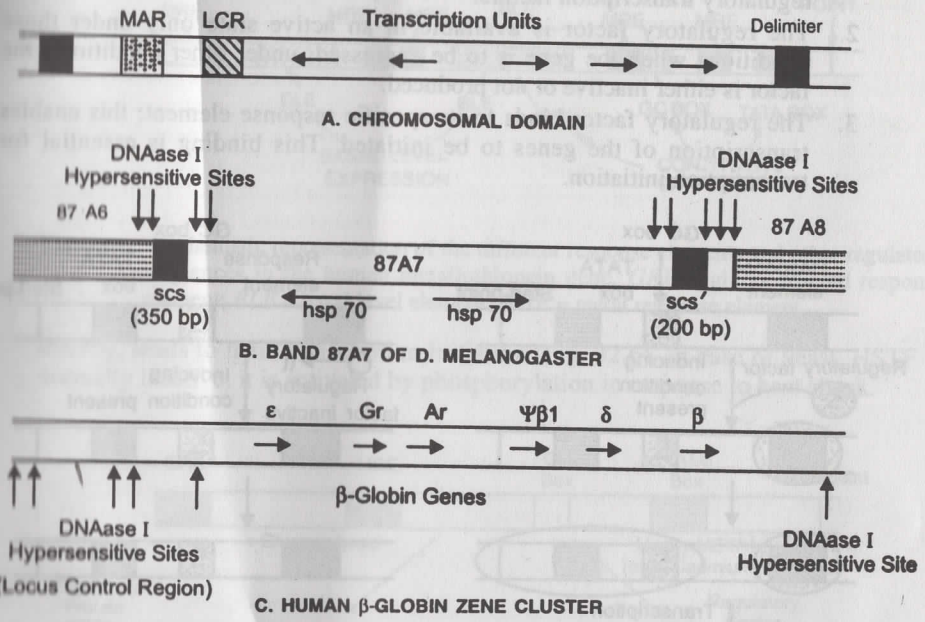


Fig. 32.2. Organization of eukaryotic chromosomal domains. A. A general model of a domain. B. Schematic representation of the domain in band *87A7* of *D. melanogaster* (scs, specialized chromatin structure). C. Human β-globin gene cluster.

4. Presumably these sites result from binding of specific regulatory proteins, *e.g.,* transcription factors, to DNA.
5. Their formation may involve histone acetylation.
6. The mechanism that regulates their formation is not known.

32.5.1. Regulation of Transcription Initiation

The genes located in active regions of chromatin are available for transcription. But the actual transcription of these gene may be regulated at the level of transcription initiation; this is likely to be the most common mechanism of regulation of eukaryotic genes. *In eukaryotes, transcription can begin only after transcription factors bind to promoter DNA, which enables RNA polymerase to bind the promoter.* Some of the transcription factors are tissue-specific; these and many other factors are involved in regulation of transcription initiation and tare called ***regulatory transcription factors***. The regulation of transcription initiation is predominantly positive, but negative regulation is also found.

32.5.1.1. Positive Regulation

In general initiation of transcription is under positive control. The general features of such a regulation are summarised below (Fig. 32.3).

1. The promoter/enhancer regions of genes subject to such regulation contain an element, called ***response element***, that serves as binding site for specific regulatory transcription factors.
2. The regulatory factor is available in an active state only under those conditions when the gene is to be expressed; under other conditions, the factor is either inactive or not produced.
3. The regulatory factor binds to its specific response element; this enables transcription of the genes to be initiated. This binding is essential for transcription initiation.

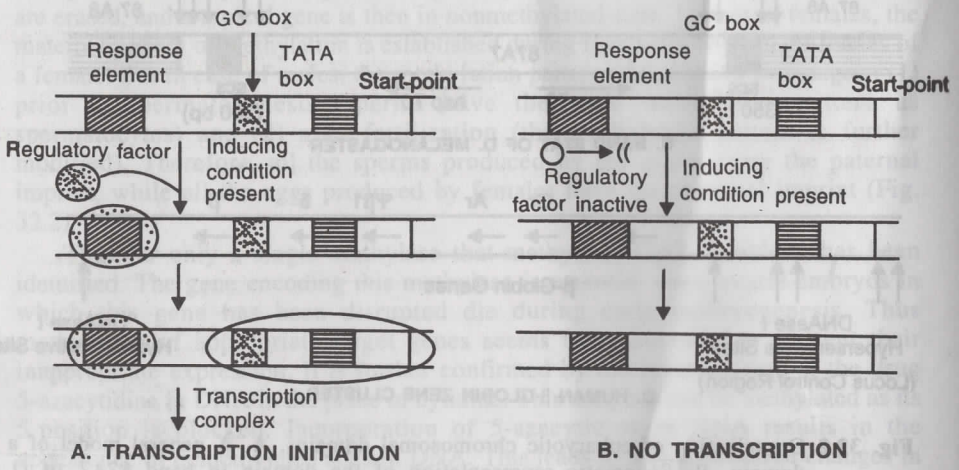


Fig. 32.3. A schematic representation of the regulation of transcription initiation in eukaryotes. (A) When inducing condition is present, the regulatory protein is in an active state; it binds the promoter and transcription initiation occurs. (B) In the absence of the inducing condition, the regulatory factor is not in active state; it can not bind the response element and transcription initiation does not take place.

This kind of gene regulation is ***inducible positive regulation*** similar to that of 'catabolite sensitive' operons of *E. coli*.

Promoters/enhancer of many genes contain more than one type of response elements. For example, the human metallothionein gene contains the following response elements: (1) GRE (gluco-corticoid response element), (2) BLE (basal level element), (3) MRE (metal response element) and (4) TRE (TPA response element) (Fig. 32.4). In such cases, binding of the appropriate regulatory transcription factor to any one of the response elements is able to activate transcription initiation (Fig. 32.5).

Often several genes may respond to the same inducing condition. All such genes are controlled by the same regulatory factor, since they contain the same response element in their promoters/enhancers. For examples, heat shock induces tanscription of ~ 20 genes in a wide variety of eukaryotes. In *Drosophila*, all the heat shock genes contain in their regulatory regions (promoters/enhancers) multiple copies of the heat shock element (HSE). HSE serves as the binding site for heat shock transcription factor (HSTF). Active HSTF binds to HSE of all heat shock genes and,

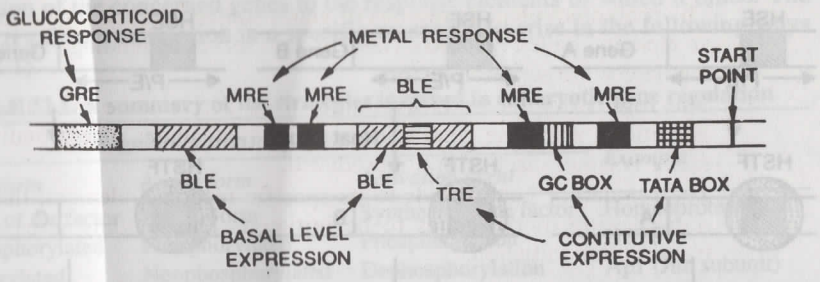


Fig. 32.4. A schematic representation of the different response elements and other regulatory sequences in the human metallothionein gene. *GRE* = glucocorticoid response element; *BLE* = basal level element; *MRE* = metal response element.

thereby, leads to their transcription initiation (Fig. 32.6). In case of yeast, HSTF is normally inactive; it is activated by phosphorylation in response to heat shock.

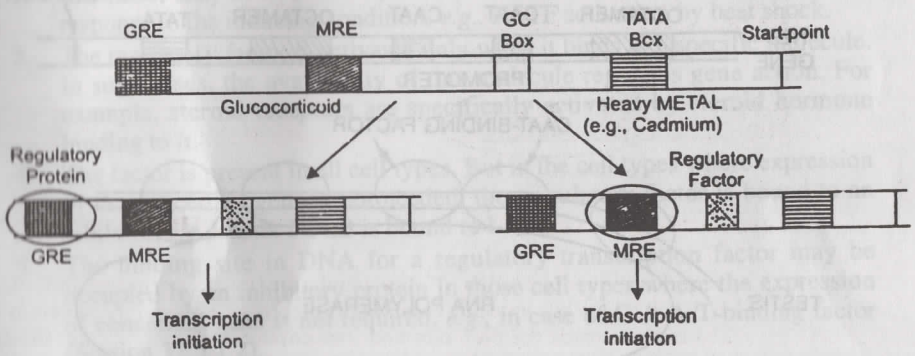


Fig. 32.5. A schematic representation of regulation of transcription initiation at a gene subject to induction by two different stimuli. Transcription is initiated by binding of either *GRE* or *MRE* by the appropriate regulatory factor. GRE, glucocorticoid response element; *MRE*, metal response element.

32.5.1.2. Negative Regulation

A good example of negative regulation is found in case of a gene encoding histone H2B in sea urchin; this gene is expressed only during spermatogenesis. The promoter of this H2B gene has two CAAT boxes; the CAAT–binding factor must bind these two boxes for transcription to be initiated (Fig. 32.7A). But in tissues other than testis, *e.g.*, in embryonic tissues, these CAAT boxes are occupied by molecules of a CAAT- displacement factor. As a consequence, the CAAT-binding factor is unable to bind the CAAT boxes, and transcription initiation can not take place (Fig. 32.7B). In testis, however, the CAAT-displacement factor does not occupy the CAAT boxes; this enables CAAT-binding factor to bind these boxes and enable transcription initiation.

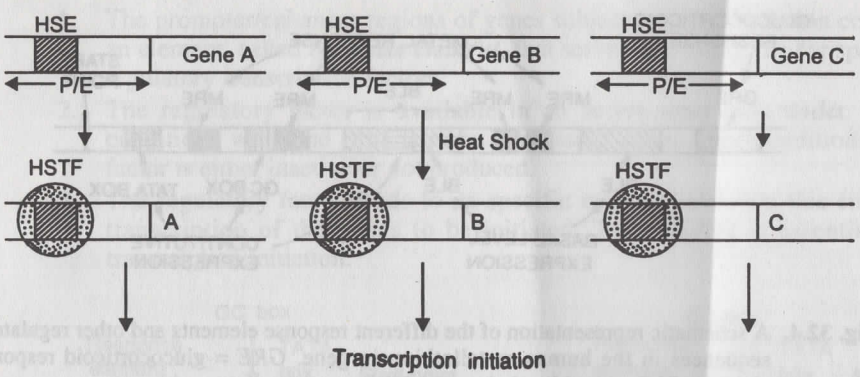


Fig. 32.6. A schematic representation of regulation of transcription initiation of several genes that respond to the same inducing condition. HSE, heat shcok element, HSTF, heat shock transcription factor; P/E, promoter/enhancer.

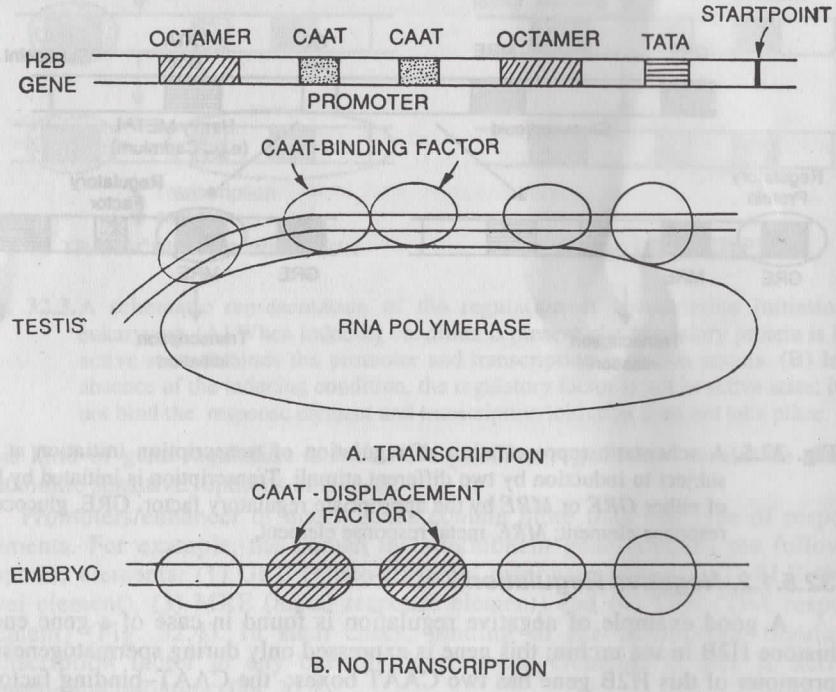


Fig. 32.7. Regulation of H2B gene. (A) In testis, the CAAT box is free so that the CAAT-binding factor binds to the box and transcription is initiated. (B) In embryo, however, the CAAT-box is occupied by the CAAT-displacement factor; as a result, the transcription can not be initiated.

32.5.2. The Basis of Regulation

Clearly, each regulatory protein must perform two functions: (1) it recognizes the specific response element and binds to them, and (2) it allows initiation of

transcription of the concerned genes to the response elements of which it binds. The ability to regulate transcription in a specific manner may arise in the following ways (Table 32.1).

TABLE 32.1. A summary of the strategies involved in eukaryotic gene regulation

Regulatory transcription factor			*Example*
Inactive form	*Active form*	*Activation event*	
Absence of the factor	Native form	Synthesis of the factor	Homeoproteins
Nonphosphorylated	Phosphorylated	Phosphorylation	HSTF
Phosphorylated	Nonphosphorylated	Dephosphorylation	Ap1 (Jun subunit)
Ligand-free	Ligand-bound	Ligand binding	Steroid receptors
Inhibitor-bound	Free from inhibitor	Release by inhibitor	NF-κB
Associated with inactive partner	Associated with active partner	Change of partner	HlH proteins (MyoD/ID)

1. The regulatory factor may be produced in certain tissues only (***tissue-specific factors***), *e.g.*, many regulatory factors having homeobox.
2. The factor may be activated by a modification, such as, phosphorylation, in response to the inducing condition, *e.g.*, HSTF activation by heat shock.
3. The regulatory factor is activated only when it binds to a specific molecule. In such cases, the availability of this molecule regulates gene action. For example, steroid receptors are specifically activated by steroid hormone binding to it.
4. The factor is present in all cell types. But in the cell types where expression of the concerned gene is not needed, the regulatory factor is bound to an inhibitor, *e.g.*, factor NFκB is bound to I-κB.
5. The binding site in DNA for a regulatory transcription factor may be occupied by an inhibitory protein in those cell types where the expression of concerned gene is not required, *e.g.*, in case of C A A T-binding factor (Section 32.5.1.2).

32.5.3. Mechanism of Action of Regulatory Factors

Regulatory factors have two main functional domains: (1) a ***DNA binding domain*** to enable them to bind to the specific response elements, and (2) an ***activation domain*** that interacts with or binds to the other components of transcription apparatus. According to an attractive model, the DNA binding and activation domains of regulatory factors are located in separate parts of the molecule; the two domains are connected together by a flexible connector (Fig. 32.8). The function of DNA-binding domain is to bind to the specific sequence in DNA is such a way that the activation domain is brought close to the startpoint. The flexible connector domain allows the activation domain to find its target protein, irrespective of the site in the promoter occupied by the DNA binding domain.

32.5.4. Dosage Compensation

Sex-linked genes that are not connected with sex determination are present in two doses in the mammalian and *Drosophila* females (because they are XX), while males (XY) contain a single dose of each. However, the intensity of phenotypes, the amounts of concerned enzymes, and even the amount of RNA produced by the

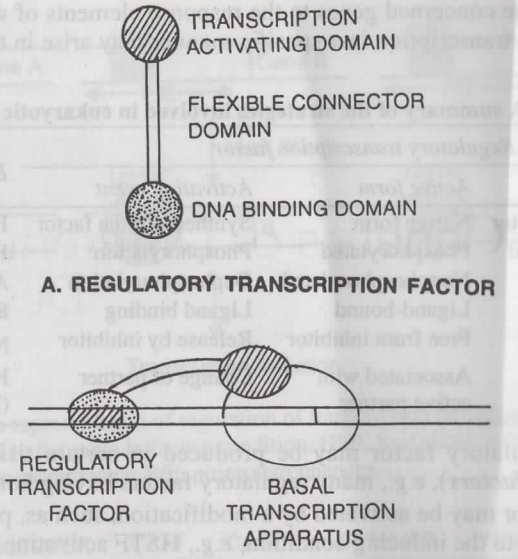


Fig. 32.8. Functional domains of a regulatory transcription factor and its interaction with the basal transcription apparatus.

X-linked genes in the females are equal to those in the males. Clearly, the expression of X-linked genes is regulated in such a manner that the level of their expression is comparable in both males and females; this is called ***dosage compensation***.

Dosage compensation occurs only in those species where male is the heterogametic sex. There are two basic mechanisms of dosage compensation: (1) inactivation of one X chromosome in females, and (2) hyperactivity of the single X chromosome in males. These aspects have been considered in some detail in Chapter 20 (Section).

QUESTIONS

1. Explain the mechanism of transcription regulation in eukaryotes.
2. Explain the regulation of gene action in eukaryotes under the following heads: (*i*) paramutation, and (*ii*) genetic imprinting.
3. What is position effect ? Explain the types, causes and the relevance of position effects in eukaryotes.
4. Write short notes on:
   (*i*) active chromatin, (*ii*) response elements,
   (*iii*) regulatory transcription factors, (*iv*) active chromatin,
   (*v*) paramutation, (*vi*) genetic imprinting,
   (*vii*) position effect.

33

Genetics of Differentiation and Development

33.1. INTRODUCTION

All multicellular organisms begin their life-cycle as single cells from which their various tissues and organs develop through differentiation. Thus ***development*** consists of the formation of different types of tissues/organs/cells, while ***differentiation*** denotes a permanent or irreversible change in the function of cells (relative to the single-celled zygote), which is often accompanied with changes in their structure. For example, cells derived from the same zygote may develop, among several others, into either skin cells, nerve cells or red blood cells; these cell types differ both in their structure as well as function (the type of gene products being synthesized). Differentiation is the most apparent in multicellular organisms, but it also occurs in unicellular organisms, *e.g.*, bacteria and viruses. *Development in both prokaryotes and eukaryotes involves the production of proteins that generate in them specific structural and functional phenotypes.*

33.2. DIFFERENTIATION IN MULTICELLULAR ORGANISMS

In multicellular organisms, the vast variety of cell types are derived through differentiation from the descendants of a single cell, the zygote. The different cell types have highly divergent morphology, and often are highly specialized for carrying out only a few specific metabolic functions. For example, more than 90% of the protein synthesized in red blood cells (RBCs) during the peak of their maximal activity is haemoglobin, only nerve cells synthesize neurotransmitters, etc. Some of the important facts known about differentiation are summarised below.

1. It was proposed by Weismann and Roux that different cell types differed in their gene content and that each cell type contains only those genes that are needed for its function. But *cells do not appear to lose any genetic material during differentiation, except in a few notable cases.* RBCs extrude their nuclei during the last stages of differentiation, and lymphocytes lose, during differentiation, those segments of antibody genes, which are not needed in a particular cell type.
2. *In any differentiated cell type, only a small proportion (less than 10%) of the genome is transcribed to yield mRNA, e.g.,* 2-5 per cent in mice liver cells, 8% in brain cells of the toad *Xenopus,* only 1% in *Xenopus* oocytes, etc.
3. *In different cell types, different sets of genes are transcribed.* The RNA

sequences present in the different cell types may differ from each other by 10-100%. For example, the mRNAs from *Xenopus laevis* oocytes and blastula (a stage in embryo development) embryos are entirely different; those from mouse liver, spleen and kidney cells differ from each other in sequence composition from about 15-70%; etc.

33.3. A GENERALIZED SCHEME FOR DIFFERENTIATION

In multicellular organisms, different body parts differ from each other in structure and function; they also occupy uniformly specific positions in all the individuals of a species. This uniformity of development in the different individuals lies in the fact that the sequence of regulatory events is ordinarily identical in all the individuals of a species. In simple words, *the pattern of development in a group of cells depends on the kinds and the sequence of their genes in active state. The stimuli for gene action may emanate from factors that control transcription or translation, or those that exercise localized control on the activities of such factors.* Most of the genes that regulate development encode regulators of transcription. These genes act in a hierarchical manner in which a group of such factors activates the transcription of genes encoding the members of the next group, and so on. These factors also activate the transcription of those genes whose products are responsible for organ formation; these genes encode kinases, cytoskeletal elements, secreted proteins and transmembrane receptors.

Egg fertilization initiates intense protein synthesis, very rapid DNA replication and rapid cell division. *Protein synthesis in embryos until the gastrula stage (a stage of embryo development) is based on mRNA molecules already present in the egg; new mRNA is transcribed only when embryos reach this stage.* There is an intense RNA synthesis during meiosis in oocytes; in vertebrates, this synthesis occurs on lampbrush chromosomes (see, Chapter 3). Huge amounts of rRNA and mRNAs are synthesized to support protein synthesis until the gastrula stage.

The distribution of different mRNAs within an egg cell is uneven; specific mRNAs become localized in specific areas of the egg. Thus, at least in most vertebrates, the genetic programmes controlling the early development (upto about the blastula stage of embryo) are established during oogenesis; these programmes are present in the egg in the form of localization of specific mRNAs. Thus the initiation of developmental pattern begins with the egg itself due to an uneven distribution of various molecules in the egg cytoplasm; these molecules may be factors concerned with regulation of transcription, or translation, or activation of such factors. Once a group of cells takes off on a developmental path due to these forces, it may produce some inducers, which would form a gradient. *Specific groups of genes would be activated in cells due to these gradients and their interactions, depending upon the position of cells in the embryo. Since these gradients occur in a preprogrammed manner, the sequences and the end-products of developmental events generated by them in different individuals of the same species will be quite uniform.*

The asymmetrical distribution of factors responsible for early development is typified by *D. melanogaster.* But in case of mammals, the egg itself is homogeneous, and the asymmetry is acquired during the initial division cycles. In both the cases, development involves differential gene expression mediated by

transcription factors, which include, a variety of proteins. The important events occurring during early development are explained using *Drosophila* as an example.

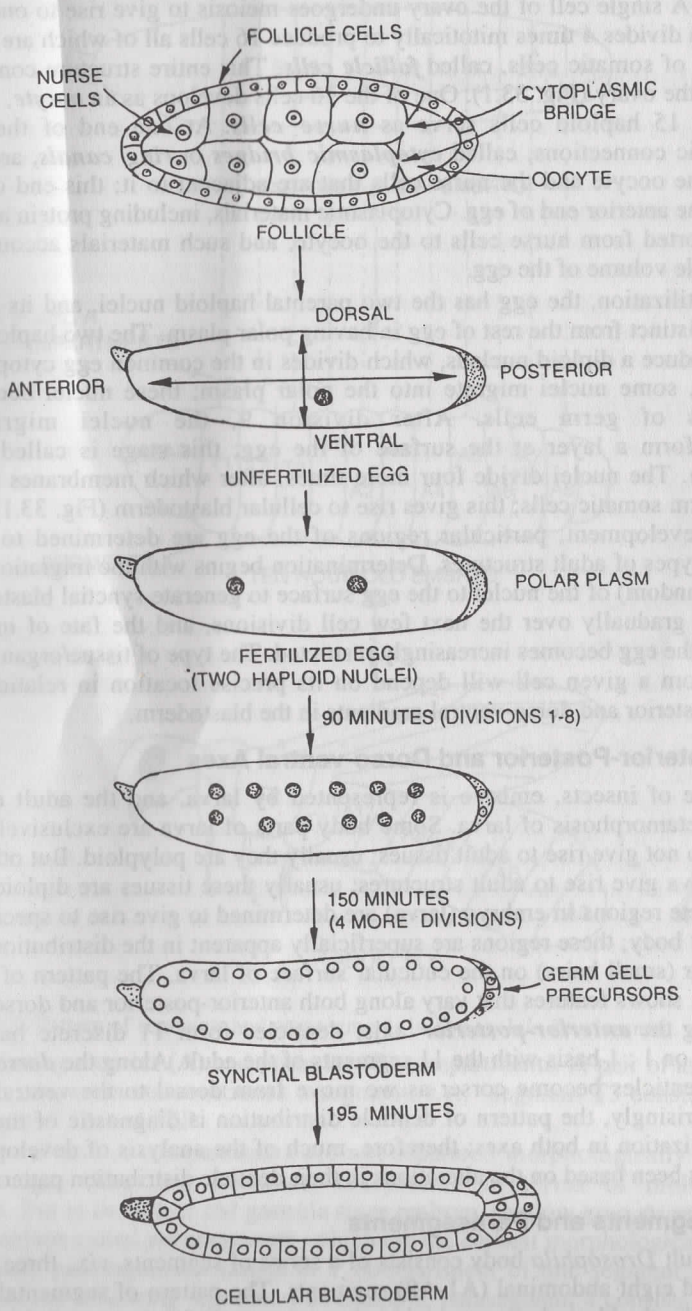


Fig. 33.1. Early development in *Drosophila melanogaster.*

33.4. EARLY DEVELOPMENT IN *DROSOPHILA*

In *Drosophila,* the asymmetry in molecular distribution is established during oogenisis. A single cell of the ovary undergoes meiosis to give rise to one haploid cell, which divides 4 times mitotically to produce 16 cells all of which are enclosed by a layer of somatic cells, called ***follicle cells***. This entire structure constitutes a ***follicle*** in the ovary (Fig. 33.1). One of the 16 cells develops as an ***oocyte***, while the remaining 15 haploid cells serve as ***nurse cells.*** At one end of the oocyte, cytoplasmic connections, called ***cytoplasmic bridges*** or ***ring canals,*** are formed between the oocyte and the nurse cells that are adjacent to it; this end of oocyte becomes the anterior end of egg. Cytoplasmic materials, including protein and RNA, are transported from nurse cells to the oocyte, and such materials accounts for a considerable volume of the egg.

At fertilization, the egg has the two parental haploid nuclei, and its posterior region is distinct from the rest of egg in having polar plasm. The two haploid nuclei fuse to produce a diploid nucleus, which divides in the common egg cytoplasm. At division 7, some nuclei migrate into the polar plasm; these nuclei become the precursors of germ cells. After division 9, the nuclei migrate and divide to form a layer at the surface of the egg; this stage is called ***synctial blastoderm***. The nuclei divide four more times, after which membranes surround them to form somatic cells; this gives rise to cellular blastoderm (Fig. 33.1). At this point in development, particular regions of the egg are determined to become particular types of adult structures. Determination begins with the migration (which occurs at random) of the nuclei to the egg surface to generate synctial blastoderm; it progresses gradually over the next few cell divisions, and the fate of individual regions of the egg becomes increasingly restricted. The type of tissue/organ that will develop from a given cell will depend on its precise location in relation to the anterior-posterior and dorso-ventral gradients in the blastoderm.

33.4.1. Anterior-Posterior and Dorso-ventral Axes

In case of insects, embryo is represented by larva, and the adult develops through metamorphosis of larva. Some body parts of larva are exclusively larval, and they do not give rise to adult tissues; usually they are polyploid. But other body parts of larva give rise to adult structures; usually these tissues are diploid. In any case, discrete regions in embryo (larva) are determined to give rise to specific parts of the adult body; these regions are superficially apparent in the distribution pattern of ***denticles*** (small hairs) on the cuticular surface of larva. The pattern of denticle distribution shows features that vary along both anterior-posterior and dorso-ventral axes. Along the ***anterior-posterior*** axis, denticles form 11 discrete bands that correspond on 1 : 1 basis with the 11 segments of the adult. Along the ***dorso-ventral axis***, the denticles become corser as we move from dorsal to the ventral side of larva. Surprisingly, the pattern of denticle distribution is diagnostic of the overall body organization in both axes; therefore, much of the analysis of development in mutants has been based on the alterations in their denticle distribution pattern.

33.4.2. Segments and Parasegments

The adult *Drosophila* body consists of a series of segments, *viz.*, three thoracic (T1-T3) and eight abdominal (A1-A8) segments. The pattern of segmental units is established by the blastoderm stage when the main mass of embryo is divided into a

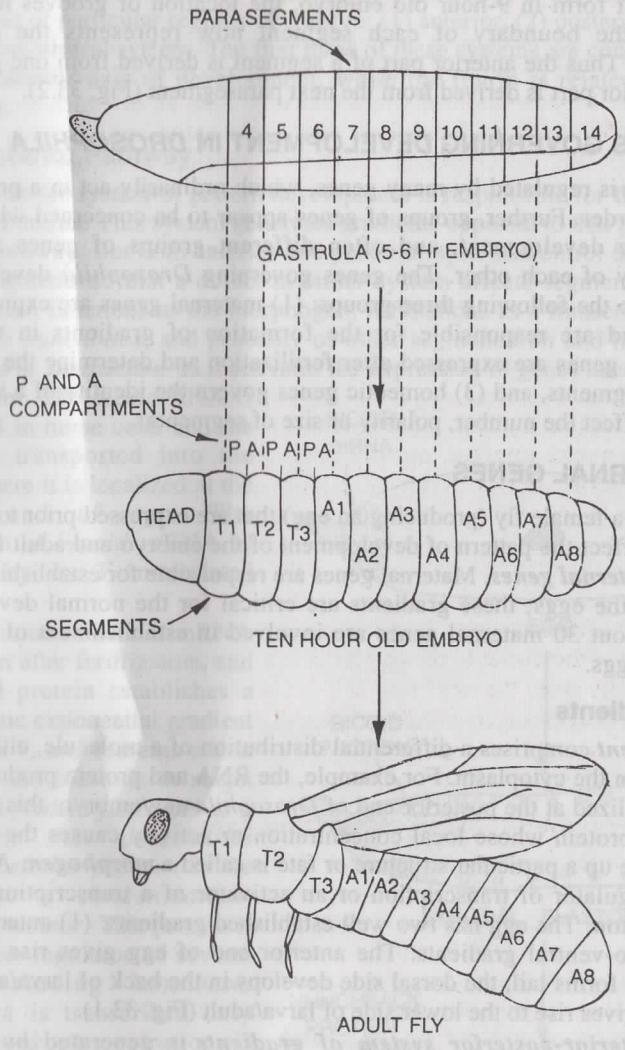


Fig. 33.2. Pattern of segments and parasegments in *Drosophila* development.

series of alternating anterior (A) and posterior (P) compartments. A pair of anterior and posterior compartments constitutes a segment, *e.g.,* segment T3 consists the compartment pair T3A and T3P.

In fact, the P and A compartments are not separated morphologically at the blastoderm stage; they are, however, discernible in terms of molecular differentiation. But in the 5-6 hr old gastrula stage embryo, shallow grooves separate the adjacent regions called ***parasegments***, which are the earliest morphological units of embryo. Each parasegment consists of P compartment of one segment and A compartment of the following segment. For example, parasegment 4 consists of the P compartment of segment T1 and the A compartment of segment T2. When

segments first form in 9-hour old embryo, the location of grooves moves to the extent that the boundary of each segment now represents the centre of a parasegment. Thus the anterior part of a segment is derived from one parasegment and its posterior part is derived from the next parasegment (Fig. 33.2).

33.5. GENES GOVERNING DEVELOPMENT IN *DROSOPHILA*

Development is regulated by many genes, which ordinarily act in a predetermined and precise order. Further, groups of genes appear to be concerned with a specific aspect of the development, and often different groups of genes seem to act independently of each other. The genes governing *Drosophila* development are classified into the following three groups: (1) maternal genes are expressed during oogenesis and are responsible for the formation of gradients in the egg, (2) segmentation genes are expressed after fertilization and determine the number and polarity of segments, and (3) homeotic genes govern the identity of a segment (but they do not affect the number, polarity or size of segments).

33.6. MATERNAL GENES

The genes of a female fly (producing an egg) that are expressed prior to fertilization and, which affect the pattern of development of the embryo and adult from the egg are called ***maternal genes***. Maternal genes are responsible for establishing the initial gradients in the eggs; these gradients are critical for the normal development of embryos. About 30 maternal genes are involved in establishment of gradients in *Drosophila* eggs.

33.6.1. Gradients

A ***gradient*** comprises a differential distribution of a molecule, either a protein or an RNA, in the cytoplasm. For example, the RNA and protein products of *nanos* gene are localized at the posterior end of *Drosophila* egg/embryo; this constitutes a gradient. A protein whose local concentration or activity causes the surrounding region to take up a particular structure or fate is called a ***morphogen***. A morphogen may be a regulator of transcription or an activator of a transcription factor in a localized region. The egg has two well established gradients: (1) anterior-posterior and (2) dorso-ventral gradients. The anterior end of egg gives rise to head, the posterior end forms tail, the dorsal side develops in the back of larva/adult, and the ventral side gives rise to the lower side of larva/adult (Fig. 33.1).

The ***anterior-posterior system of gradients*** is generated by differential distribution of molecules along the anterior-posterior axis of the egg; this gradient is established soon after fertilization. It controls the positional information along the larva, *e.g.*, location of head, thorax, abdomen etc. The ***dorso-ventral gradient*** system, on the other hand, comprises differential activation of molecules leading to differential distribution of some other molecules (Section 33.6.4); it is generated somewhat later than the anterior-posterior gradient. This system regulates tissue differentiation, *i.e.*, specification of distinct embryonic tissues, including mesoderm, neuroectoderm and dorsal ectoderm.

The two body axes are established independently. For example mutations that affect polarity lead to the development of anterior structures in posterior regions or those of dorsal structures in the ventral regions. But mutations affecting one axis, say, anterior-posterior axis, do not affect the other axis, that is, dorso-ventral axis.

The maternal genes constitute the following four pathways or systems that affect the development of particular regions of embryo: (1) anterior, (2) posterior, (3) terminal and (4) dorso-ventral system. The first three of these systems are concerned with the anterior-posterior axis of development, while the fourth is related to the dorso-ventral axis.

33.6.2. Anterior Pathway

The ***anterior system*** of pattern development is responsible for the development of head and thorax. This system generates gradients opposite to that produced by the posterior pathway. But both anterior and posterior systems control the development of body segments so that a defect in either system affects segment development. Bicoid protein functions as the morphogen; it functions as a transcription regulator of the gene *hunchback*, and probably of other segmentation and homeotic genes. Thus *bicoid* is ***instructive*** as it activates the expression of genes needed for anterior development. The *bicoid* RNA is transcribed in nurse cells and the mRNA is transported into the oocyte, where it is localized at the anterior end of egg. The localization of *bicoid* mRNA is specified by its 3′-untranslated region.

Translation of *bicoid* mRNA begins soon after fertilization, and the Bicoid protein establishes a characteristic exponential gradient by division 7, and remains stable until after the blastoderm stage. *The Bicoid protein functions as a morphogen that determines anterior-posterior position in the embryo in a concentration-dependent manner.* This protein is transcription regulator of several genes, including *hunchback; hunchback* is transcribed only when Bicoid concentration is above a certain threshold level. In this way, the quantitative differences in the concentration of a morphogen are converted into qualitative differences in cell function and, ultimately, structures during embryonic development. As a result, *hunchback* product forms a band in the anterior part of the embryo (Fig. 33.3).

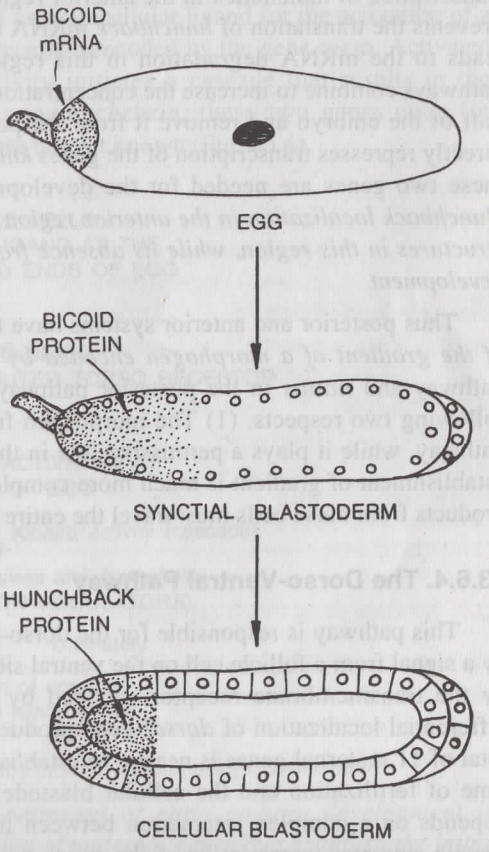


Fig. 33.3. The anterior development pathway of early development of *Drosophila* embryos.

33.6.3. The Posterior System

This system is responsible for the development of the segments of abdomen. A large number of events cause the localization of *nanos* mRNA at the posterior end of the egg, *i.e.*, at the pole the Nanos protein functions as the morphogen. Nanos protein leads to a localized repression of the translation of *hunchback* mRNA, thereby enforcing the anterior localization of Hunchback protein. The posterior localization of the *nanos* mRNA depends upon sequences in the 3′-untranslated region of the mRNA. Although *nanos* mRNA is localized at the posterior pole of egg, Nanos protein diffuses to form a gradient that extends along the abdominal region.

Gene *hunchback* is transcribed during oogenesis and the mRNA is uniformly distributed in the egg cytoplasm. After fertilization, Bicoid gradient activates transcription of *hunchback* in the anterior region. At the same time, Nanos gradient prevents the translation of *hunchback* mRNA in the posterior region; this inhibition leads to the mRNA degradation in this region. Thus the posterior and anterior pathways combine to increase the concentration of *hunchback* mRNA in the anterior half of the embryo and remove it from the posterior half. The Hunchback protein directly represses transcription of the genes *knirps* and (probably) *giant;* products of these two genes are needed for the development of abdominal structures. *Thus Hunchback localization in the anterior region prevents the formation of abdominal structures in this region, while its absence from the posterior region ensures their development.*

Thus posterior and anterior systems have the same basic feature: *establishment of the gradient of a morphogen encoded by a maternal gene* (*bicoid* in anterior pathway and *nanos* in the posterior pathway. But the two systems differ in the following two respects. (1) The morphogen functions instructively in the anterior pathway, while it plays a permissive role in the posterior pathway. (2) Further, the establishment of gradient is much more complex in the posterior pathway since the products from nurse cells must travel the entire length of egg to reach the pole.

33.6.4. The Dorso-Ventral Pathway

This pathway is responsible for the dorso-ventral development, and is initiated by a signal from a follicle cell on the ventral side of the egg. The signal is perceived by the transmembrane receptor encoded by *Toll,* whose activation leads to a differential localization of *dorsal* gene product, which is a transcription factor. A total of 11 maternal genes is needed to establish the dorso-ventral axis between the time of fertilization and the cellular blastoderm. The dorso-ventral development depends on a complex interaction between the oocyte and follicle cells. Dorsal protein activates transcription of the genes *twist* and *snail*; these gene products are required for the development of ventral structures. At the same time, Dorsal represses the genes *dpp* and *zen,* which are required for the development of dorsal structures. As a result, *dpp* and *zen* are expressed only in the 40% most dorsal of the

embryo, which develops the dorsal structures. The protein Dpp is, in effect, the morphogen that induces the formation of dorsal structures.

The principle of dorso-ventral development in flies, amphibians and vetebrates is comparable, but there is one remarkable difference: the neural structures, including the central nervous system, develop on the ventral side of flies but on the dorsal side of the vertebrates. Consistent with this, the signaling pathway on the dorsal side of flies is comparable to that on the ventral side of vertebrate embryos, where nonneural structures develop.

33.6.5. The Terminal Pathway

The terminal system is responsible for the development of the specialized structures at the unsegmented ends of the egg, *viz.,* acron at head and the telson at the tail. The initial events that create asymmetry occur in the follicle cells; these events lead to a localized production of an extracellular ligand for the activation of a uniformly distributed transmembrane receptor encoded by the gene *torso.* Activated Torso has a kinase activity; this activity initiates a cascade that results in the localized expression of genes *tailless* and *huckebein;* these two genes code for transcription regulators whose targets are not yet known (Fig. 33.4).

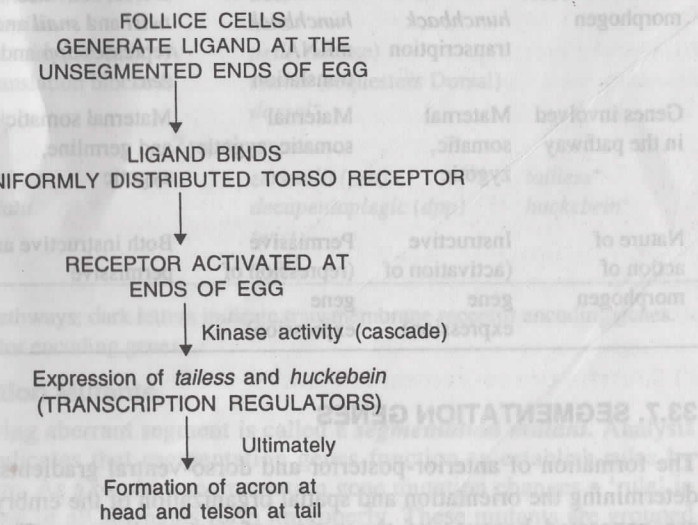


Fig. 33.4. The terminal pathway of early embryo development in *Drosophila.*

Thus the patterns of the four pathways of early embryo development in Drosophila are established by the action of maternal genes that establish the initial asymmetry in the egg and early embryo. The various features of the four pathways have been summarised in Table 33.1; the four pathways differ in respect of several important features. Table 33.2 lists the various genes involved in the four pathways and attempts to identify the critical functions in the pathway.

TABLE 33.1
A comparison of the important features of the four pathways of early embryo development in *Drosophila*

	Pathway			
Feature	***Anterior***	***Posterior***	***Dorso-ventral***	***Terminal***
Initial asymmetry created by	Localization of an mRNA (*bicoid*)	Localization of an mRNA (*nanos*)	Localized activation of a receptor (Toll)	Localized activation of a receptor (Torso)
Signal for asymmetry resides in	3′-region of the mRNA	3′-region of the mRNA	Localized production of ligand (Spatzle)	Localized production of ligand
Gradient established by	Uneven distribution in cytoplasm	Uneven distribution in cytoplasm	Nuclear and cytoplasmic distribution (concentration is even throughout)	—
Morphogen	Bicoid	Nanos	Dorsal (ventral side) and Dpp (dorsal side)	—
Action of morphogen	Activation of *hunchback* transcription	Prevention of *hunchback* mRNA translation	Dorsal activates *twist* and *snail* and represses *dpp* and *zen*	Expression of *tailless* and *huckebein*
Genes involved in the pathway	Maternal somatic, zygotic	Maternal somatic, zygotic	Maternal somatic and germline, zygotic	Maternal somatic and germline, zygotic
Nature of action of morphogen	Instructive (activation of gene expression)	Permissive (repression of gene expression)	Both instructive and permissive	Instructive

33.7. SEGMENTATION GENES

The formation of anterior-posterior and dorso-ventral gradients is the first step in determining the orientation and spatial organization of the embryo. These gradients are generated because of the action of maternal genes. The next step in the development is the formation of discrete regions that are determined to give rise to different body parts; this depends on a differential expression of the genes of the embryo *per se*, and these genes are called ***zygotic genes***. The zygotic genes are identified by segmentation mutants; therefore, they are also known as ***segmentation genes***. The segmentation genes are expressed after fertilization and their products form transverse bands that differentiate individual regions on the anterior posterior axis. Approximately 30 different loci are involved in segment formation.

TABLE 33.2
A summary of the various genes involved in the early embryonic development in *D. melanogaster*

Pathway			
Anterior	***Posterior***	***Dorso-ventral***	***Terminal***
		Maternal somatic genes	
		pipe	*torsolike*
		nudel	
		windbeutel	
		Maternal germline genes	
exuperantia	*spire*	*gastrulation-defective*	*trunk*
swallow	*cappuccino*	*snake*	*Nasrat*
staufen	*staufen*	*easter*	*polehole*
	oskar	*spatzle*	
	vasa		
	valois		
	tudor		
	mago nashi		
*bicoid**+	*nanos**	***Toll****	***torso****
	pumilio	*tube*	
	hunchback+ (mRNA translation blocked)	*pelle* (kinase)	
		cactus (sequesters Dorsal)	
		dorsal+	
		Zygotic genes	
hunchback	*knirps*	*zerknullt (zen)*	*tailless*+
buttonhead	*giant*	*decapentaplegic (dpp)*	*huckebein*+
orthodenticle		*twist*	
empty spiracles		*snail*	

* Morphogen in the pathways; dark letters indicate transmembrane receptor encoding genes.
+ Transcription regulator encoding genes.

33.7.1 Segmentation Mutants

A mutant showing aberrant segment is called a ***segmentation mutant.*** Analysis of these mutants indicates that segmentation genes function to establish rules by which segments form. As a result, a segmentation gene mutation changes a 'rule' in such a way that many or all segments form improperly. These mutants are grouped into 3 classes, *viz.*, (1) gap mutants, (2) pair rule mutants and (3) segment polarity mutants. These groups of genes are expressed in a successive manner, and each successive group defines an increasingly restricted regions of the embryo. Products of maternal genes regulate gap genes, those of gap genes regulate pair-rule genes, which, in turn, regulate segment polarity genes; this may be represented as follows. Maternal genes → Gap genes → Pair-rule genes → Segment polarity genes.

33.7.2. Gap Genes

Mutants in these genes have a group of several adjacent segments deleted from the final pattern. Four gap genes are involved in the formation of the major body

segments and others are concerned with head and tail structures. The important gap genes are, *hunchback, kruppel, knirps, giant* and *tailless*. The maternal genes that establish the anterior-posterior axis generate Bicoid and Nanos gradients; these gradients either activate or repress the expression of gap genes. The gap genes are transcribed following the 11th nuclear division and divide the embryo into 4 broad regions (Fig. 33.5), which are produced as follows.

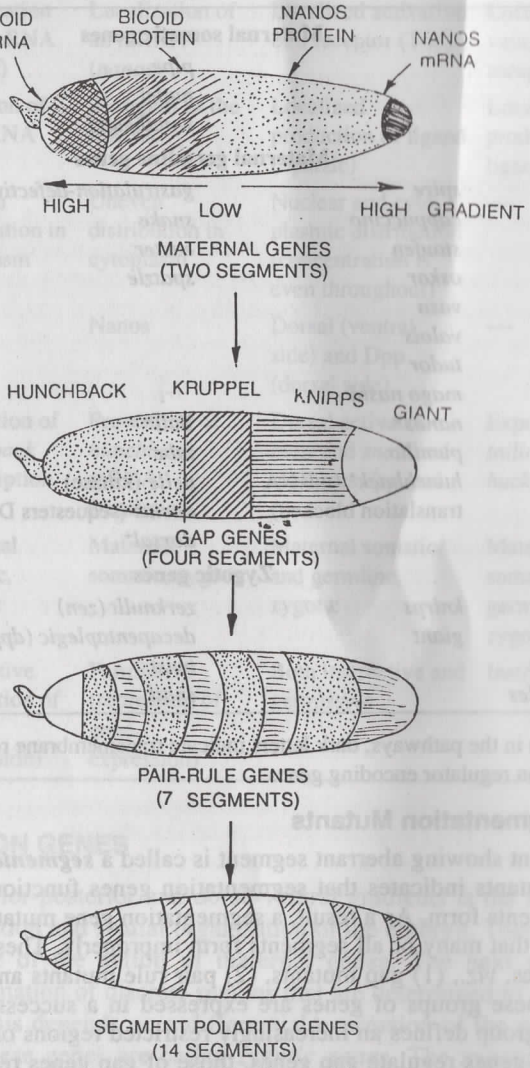


Fig. 33.5. Segmentation genes divide *Drosophila* embryos ito successively smaller segments.

1. The most anterior band consists of Hunchback protein; transcription of *hunchback* is activated at the anterior end by Bicoid protein. High levels of Hunchback repress *kruppel;* hence this region does not have Kruppel protein.

2. The next band has the Kruppel protein. Transcription of *kruppel* gene is activated by low levels of the Hunchback protein.
3. The third band is formed by Knirps protein, while the fourth band is due to the Giant protein. Transcription of *knirps* and *giant* is repressed by Hunchback in the anterior region. The Nanos protein prevents translation of *hunchback* mRNA in the posterior region, which allows the expression of *knirps* and *giant* in the posterior region.

Thus gap genes are not only regulated by the maternal genes, but they are also regulate each other. *All the gap proteins are regulators of transcription. In addition to regulating the expression of each other, they regulate the expression of the next set of segmentation genes, called pair-rule genes.*

33.7.3. Pair-Rule Genes

Mutants in ***pair-rule genes*** affect either all odd-numbered or all even-numbered segments, and in every affected segment the same part of pattern is deleted. There are 8 pair-rule genes, *viz., runt, hairy, ftz, even skipped, paired, odd paired, sloppy paired* and *odd skipped.* The earliest time of expression of these genes is at nuclear divisions 11 to 12. Thus each pair-rule protein is found in a pattern of 7 stripes representing either all the odd-numbered segments or all the even-numbered segments. Genes *hairy* and *eve* are called ***primary pair-rule genes*** because they are expressed first and their pattern of expression influences the expression of other pair-rule genes.

The expression of pair-rule genes is controlled by combination of different proteins in the local areas. This is illustrated by taking the example of the genes *eve* and *ftz* (*fushi tarazu*). Gene *eve* is expressed in the 7 odd-numbered segments, while *ftz* is expressed in the remaining 7 even-numbered segments. Initially, the bands of *eve* and *ftz* proteins are fuzzy at the blastoderm stage; they become sharper as development proceeds to gastrula stage. This may involve an autoregulatory loop, *i.e.,* expression of the gene is activated by its own product.

In general, generally distributed proteins, *e.g.,* Bicoid and/or Hunchback, are needed for activation of pair-rule genes, while proteins with more restricted distribution, such as, Giant and Kruppel, repress transcription and define the boundaries of the stripes. It may be pointed out that the hierarchy of gene control is not exclusively restricted to interactions between successive stages of control. For example, Bicoid is a maternal gene product, but it acts on a pair-rule gene *eve*.

33.7.4. Segment Polarity Genes

Mutations in these genes most often lead to a loss of a part of the P compartment of each segment, but some mutations lead to a loss of A compartment or the middle of segments. When P compartment is affected, it is replaced by a mirror image duplication of the A compartment of the same segment. So far, 16 polarity genes are known, *e.g., engrailed, wingless, gooseberry, patched, hedgehog, coatal2, fused,* etc. The expression of these genes occurs at nuclear division 13, and is regulated by the products of pair-rule genes; these genes are expressed in 14 stripes in a highly precise pattern. Segment polarity genes further subdivide each of the 14 segments (parasegments) formed by the pair-rule genes. The type of subdivision induced by a segment polarity gene is repeated in every segment. For example, *engrailed* function (Engrailed is a transcription factor) separates A and P

compartments in each segment. The functions of other segment polarity genes are not clear, but their products include secreted proteins, transmembrane proteins, kinases, cytoskeletal proteins as well as transcription factors. Obviously, at this stage, cell to cell interactions become important.

33.8. HOMEOTIC GENES

Homeotic genes act during embryogenesis and control the identity of a segment; they do not change the number, polarity or size of segments. Homeotic genes ensure the unique differentiation of each segment. Mutations in these genes cause cells of one compartment to develop the phenotype of a different compartment. Some homeotic mutations 'transform' part of a segment or an entire segment into another type of segment. Most homeotic genes are expressed in a spatially restricted manner that corresponds to parasegments. The expression of these genes depends on the signals established by segmentation genes. Many homeotic genes code for transcription factors that act upon other homeotic genes as well as on other target loci. The complex interlocking interactions among different homeotic genes are not yet fully understood. The segment polarity genes are active at the same time as the peak expression of homeotic genes.

Homeotic mutations cause one segment of the abdomen to develop as another, legs to develop in place of antennae or wings to develop in place of eyes. Homeotic genes do not create new patterns; they only change cell fates by switching the set of genes that functions in a particular compartment. Some homeotic genes are complex loci, *e.g., ANT-C* and *BX-C;* these two complex loci regulate the development of all the segmented units of adult insect body, and each complex locus contains several homeotic genes. ***Complex loci*** possess the following features: (1) rather well-spaced groups of mutations spread over a relatively large map distance, (2) each cluster of mutations corresponds roughly to an exon, (3) some pairs of mutations show complementation, while others do not, and (4) individual mutations may have different and complex morphological effects on phenotype. Complex loci have the same general organization as typical split genes; their unusual properties arise because they have an array of regulatory elements that control their expression in different parts of the embryo. The genes are extremely large, and usually there are two or more genes within a complex locus.

33.9. TISSUE-SPECIFIC DNA METHYLATION

In most plants and animals, cytosine residues are methylated (to yield 5-methyl cytosine) after DNA replication. The extent of methylation varies from species to species (about 2-7% of the cytosine residues in mammals). The methyl group of 5-methyl cytosine lies exposed in the major groove of DNA; thus it has the potential to be involved in the interaction of DNA with the various proteins. This conclusion is strongly supported by the fact that the addition or removal of a single methyl ($-CH_3$) group from the *lac* (operon) operator DNA sharply changes its affinity for the *lac* repressor.

Such a direct evidence for the involvement of methylation in gene regulation is lacking in eukaryotes, but there are three kinds of indirect evidence. (1) The degree of DNA methylation has been found to be negatively correlated with gene expression, that is, the higher the degreee of methylation the lower the extent of gene expression. (2) In some cases, DNA methylation patterns are tissue-specific, *i.e.,*

different tissues of the same organism show different patterns of methylation. (3) The base analogue 5-azacytidine, which is also used as a drug, is incorporated into DNA in place of cytosine but it can not be methylated. The incorporation of 5-azacytidine in DNA (following a treatment with the drug) has been shown to result in the expression of genes in those tissues where they normally are not expressed, *e.g.,* foetal haemoglobin (γ-haemoglobin) and, to a lesser extent the embryonic haemoglobin (ε-haemoglobin) genes in anaemic adult baboons and adult humans

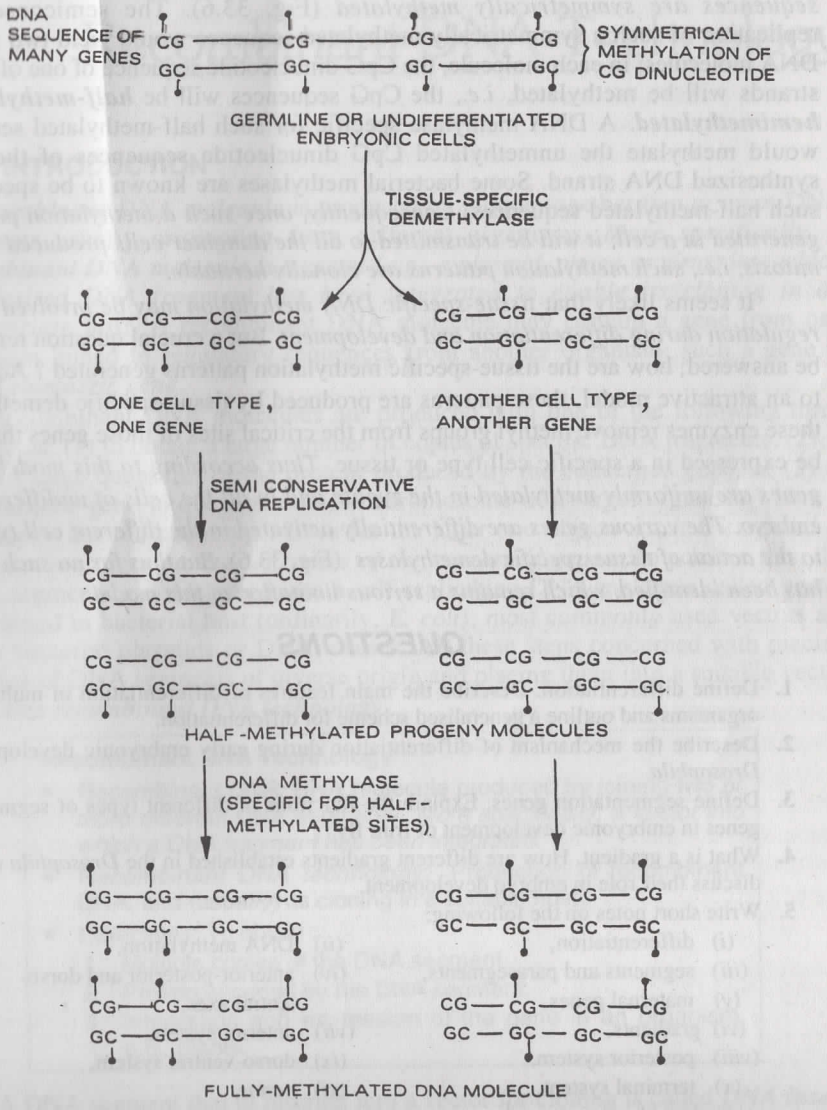


Fig. 33.6. A schematic representation of the production and maintenance of tissue-specific methylation patterns in DNA. Only CG dinucleotide sequence is shown. The solid circles attached to C residues represent 5-methyl ($-CH_3$) groups.

with severe β-thalassemia (inability to synthesize the β-chain of haemoglobin) and with sickle cell anaemia. These genes are normally not expressed in the RBCs of adults. In one of these studies, DNA in the regions of γ-haemoglobin and ε-haemoglobin genes was shown to be hypomethylated, *i.e.*, contained fewer methyl groups than normal.

In most eukaryotes, more than 90% of the methylation occurs on cytosine residues present in CpG dinucleotide sequence. Further, the CpG dinucleotide sequences are symmetrically methylated (Fig. 33.6). The semiconservative replication of such a symmetrically methylated sequence would yield two progeny DNA molecules; in each molecule, the CpG dinucleotide sequence of one of the two strands will be methylated, *i.e.*, the CpG sequences will be ***half-methylated*** or ***hemimethylated.*** A DNA methylase specific for such half-methylated sequences would methylate the unmethylated CpG dinucleotide sequences of the newly synthesized DNA strand. Some bacterial methylases are known to be specific for such half-methylated sequences. *Consequently, once such a methylation pattern is generated in a cell, it will be transmitted to all the daughter cells produced through mitosis, i.e., such methylation patterns are clonally heritable.*

It seems likely that *tissue-specific DNA methylation may be involved in gene regulation during differentiation and development.* But a crucial question remains to be answered; how are the tissue-specific methylation patterns generated ? According to an attractive model, these patterns are produced by tissue-specific demethylases; these enzymes remove methyl groups from the critical sites of those genes that are to be expressed in a specific cell type or tissue. *Thus according to this model, all the genes are uniformly methylated in the zygote and in all the cells of undifferentiated embryo. The various genes are differentially activated in the different cell types due to the action of tissue-specific demethylases* (Fig. 33.6). *But thus far no such enzyme has been identified, which remains a serious limitation of this model.*

QUESTIONS

1. Define differentiation. Describe the main features of differentiation in multicellular organisms and outline a generalised scheme for differentiation.
2. Describe the mechanism of differentiation during early embryonic development in *Drosophila.*
3. Define segmentation genes. Explain the functions of different types of segmentation genes in embryonic development of fruit fly.
4. What is a gradient. How are different gradients established in the *Drosophila* egg, and discuss their role in embryo development.
5. Write short notes on the following:
   (*i*) differentiation, (*ii*) DNA methylation, (*iii*) segments and parasegments, (*iv*) anterior-posterior and dorso-ventral axes, (*v*) maternal genes, (*vi*) gradients, (*vii*) anterior system, (*viii*) posterior system, (*ix*) dorso-ventral system, (*x*) terminal system, (*xi*) *nanos,* (*xii*) *bicoid,* (*xiii*) morphogen, (*xiv*) gap genes, (*xv*) pair-rule genes, (*xvi*) segment polarity genes, (*xvii*) homeotic genes.

34
Recombinant DNA Technology

34.1. INTRODUCTION

A ***recombinant DNA molecule*** is produced by joining together two or more DNA segments usually originating from different organisms. *More specifically, a* ***recombinant DNA molecule*** *is a vector (e.g., a plasmid, phage or virus) into which the desired DNA fragment has been integrated to enable its cloning in an appropriate host.* A recombinant DNA molecule may contain a gene from one organism joined to regulatory sequences from another organism; such a gene is called ***chimaeric gene***.

Recombinant DNA molecules are produced with one of the following three objectives: (1) to obtain a large number of copies of specific DNA fragments, (2) to recover large quantities of the protein produced by the concerned gene, or (3) to integrate the gene in question into the chromosome of a target organism where it produces a desired phenotype. Even for the latter two objectives, *it is essential to first obtain a large number of copies of the concerned genes.* To achieve this, the DNA segments are integrated into a self-replicating DNA molecule called ***vector*** and cloned in bacterial host (ordinarily, *E. coli*); most commonly used vectors are either bacterial plasmids or DNA viruses. All these steps concerned with piecing together of DNA segments of diverse origin and placing them into a suitable vector constitute ***recombinant DNA technology***.

Recombinant DNA Technology

- *Recombinant DNA:* DNA molecule produced by joining two or more different DNA segments. *More specifically; a vector into which a DNA fragment has been integrated.*
- *Recombinant DNA technology:* Production of recombinant DNA, and (usually) its cloning in a suitable host.
- *Objectives:* To obtain :
  1. Multiple copies of the DNA segment.
  2. Protein encoded by the DNA segment.
  3. Integration and expression of the gene in an organism, *e.g.,* plants.

A DNA segment that in inserted into a vector for cloning is called ***DNA insert.*** The vectors containing a DNA insert are generally called ***recombinant DNA.*** Such vectors are then introduced into a suitable organism, usually a bacterium; this organism is called ***host,*** while the process is called ***transformation***. The steps concerned with transformation of a suitable host with a recombinant DNA, and

cloning of the transformed cells are called ***DNA cloning*** or ***gene cloning***. However, often ***DNA*** or ***gene cloning*** is taken to include both the development of recombinant DNA as well as its cloning in a suitable host. Similarly, often the term ***recombinant DNA technology*** is used as a synonym for ***DNA*** or ***gene cloning*** used in the broader sense. A rather popular term for these activities is ***genetic engineering***.

Restriction Enzymes; Features

- Endonucleases (Type II) that cleave DNA either within or just outside a highly specific recognition sequence
- Over 350 different enzyme known.
- Used for DNA cleavage in recombinant DNA technology.
- *Recognition Sequence:* Specific DNA sequences for different enzymes. Their features are:
  1. Palindromes with rotational symmetry, *e.g.*, GAATTC. / CTTAAG
  2. Usually of 4, 5, or 6 bp.
  3. Predominantly G:C-rich.

34.2. RESTRICTION ENDONUCLEASES

Endonucleases are enzymes that produce internal cuts, called ***cleavage***, (both the strands are cut), in DNA molecules. A class of endonucleases cleaves DNA only within or near those sites, which have specific base sequences; such endonucleases are known as ***restriction endonucleases***, and the sites recognised by them are called ***recognition sequences*** or ***recognition sites***. The recognition sequences are different and specific for the different restriction endonucleases or restriction enzymes.

The presence of restriction enzymes was postulated by W. Arber during 1960s, while the first true restriction endonuclease was isolated in 1970. Smith, Nathans and Arber were awarded the Nobel Prize for Physiology and Medicine in 1978 for the discovery of endonucleases.

Restriction endonucleases are indispensable for DNA cloning and DNA sequencing. They serve as the tools for cutting DNA molecules at predetermined sites, which is the basic requirement for gene cloning or recombinant DNA technology. There are three distinct types of restriction endonucleases: (*i*) Type I, (*ii*) Type II and (*iii*) Type III. Of these, type II restriction endonucleases are used in recombinant DNA technology. *Type II restriction endonucleases* induce cleavage either, in most cases, within their recognition sequences (see later) or very close to them. More than 350 different type II endonucleases with over 100 different recognition sequences are known (Table 34.1).

34.2.1. Recognition Sequences

*The **recognition sequences** for Type II endonucleases form palindromes with rotational symmetry* (Table 34.1). In a ***palindrome***, the base sequence in the second half of a DNA strand is the mirror image of the sequence in its first half (Fig. 34.1).

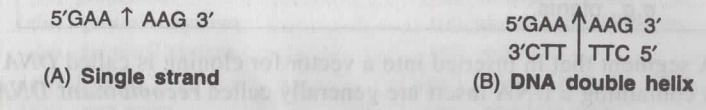


Fig. 34.1. A palindromic sequence. (A) Sequence in a single DNA strand, (B) Sequence in a DNA double helix. The arrow represents the axis of symmetry.

TABLE 34.1

Some restriction endonucleases, the organisms in which they naturally occur (source), the nucleotide sequences they recognise, and the sites of the cuts induced by them (shown by oblique lines)

Restriction endonuclease	*Source (organism and strain)*	*Recognition sequence*
*Aval**	*Anabaena variabilis* (ATCC 27892)	C/Py CG Pu G G Pu G C Py/C
Alul	*Arthrobacter luteus*	AG/CT TC/GA
*Bam*HI	*Bacillus amyloliquefaciens* H	G/G AT CC CCTA G/G
*Eco*RI	*Escherichia coli* Ry13	G/AA TTC C TT AA/G
*Eco*RII**	*E. coli* R245	/CCA (T) GG GGT (A) CC/
*Bgl*II	*Bacillus globigli*	A/G ATC T T C TAG/A
*Hind*II	*Haemophilus influenzae* Rd	GT Py/Pu AC CA Pu/Py TG
*Hind*III	*H. influence* Rd	A/AGCT T T TCGA/A
*Hin*PI	*H. influence* P_1	G/G CC CC G/G
*Hpa*I	*H. parainfluenzae*	GTT/AAC CAA/TTG
*Hpa*II	*H. parainfluenzae*	C/C GG GG C/C
*Hin*fI*	*H. influenzae* Rf	G/ANT C C TNA/G CTNA/G
*Nla*III	*Neisseria lactamica*	CATG/ /GTAC
*Pst*I	*Providencia stuartii*	CTGCA/G G/A CGTC
*Sau*3A	*Staphylococcus aureus* 3A	/GATC CTAG/
*Taq*I	*Thermos aquaticus* YTI	T/CG A AG C/T AGC/T

*Pu, either of the two purines (adenine or guanine)
Py, either of the two pyrimidines (thymine or cytosine)
N, any of the four bases (adenine, guanine, thymine or cytosine)
**The base given within parenthesis, *e.g.*, (T) may occur in place of the preceding base, *e.g.*, A.

But in a ***palindrome with rotational symmetry***, the base sequence in the first half of one strand of a DNA double helix is the mirror image of the second half of its complementary strand (Fig. 34.2). Thus in such palindromes, the base sequence in both the strands of a DNA duplex reads the same when read from the same end (either 5′- or 3′-end of both the strands).

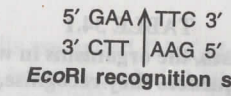


Fig. 34.2. A pallindrome with rotational symmetry. The arrow represents the axis of symmetry.

Most of the type II restriction endoucleases have recognition sequences of 4, 5 or 6 bp (base pairs), which are predominantly GC-rich.

Restriction Enzymes: Cleavage
- Usually within the recognition sequence.
- In some cases, just outside the recognition sequence.
- Most enzymes produce protruding ends, which are cohesive.
- Many enzymes generate blunt ends or flush ends.
- Cohesive ends facilitate the production of recombinant DNA.

34.2.2. Cleavage Patterns

Most type II restriction endonucleases cleave the DNA molecule within their specific recognition sequences, but some produce cuts immediately outside the target seqeunce, *e.g.*, *Nla*III, *Sau*3A, etc. (Table 34.1). These cuts are either (1) staggered or (2) even, depending on the enzyme. In ***staggered cuts,*** the two strands of a DNA double helix are cleaved at different locations; this generates ***protruding*** (3′ or 5′) ***ends*** (Fig. 34.3), *i.e.*, one strand of the double helix extends some bases beyond the other. Due to the palindromic (symmetrical) nature of the target sites, the two protruding ends generated by such a cleavage by a given enzyme have

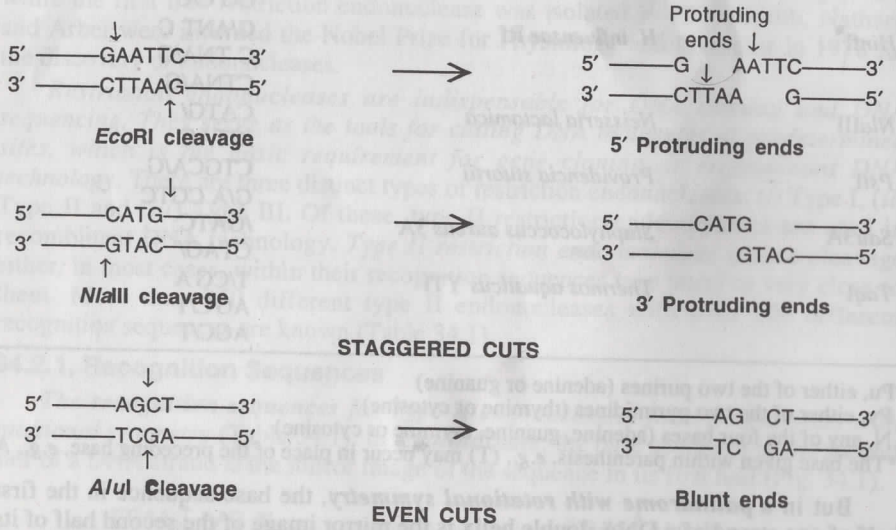


Fig. 34.3. DNA cleavage by restriction endonucleases. Staggered cuts: 5′ producing (*Eco*RI) and 3′ (*Nla*III) protruding ends, and event cuts (*Alu*I) producing bunt ends. The vertical arrows indicates the site of cut in DNA strand.

complementary base sequences. As a result, they readily pair with each other; such ends are called ***cohesive*** or ***sticky ends***. Therefore, when fragments generated by a single restriction enzyme from different DNAs are mixed, they join together due to their sticky ends (Fig. 34.4). This property of the restriction enzymes is of great value for the construction of recombinant DNAs. When both the strands of a DNA molecule are cut at the same site, ***blunt*** or ***flush*** ends are produced in which the two strands end at the same point (Fig. 34.3).

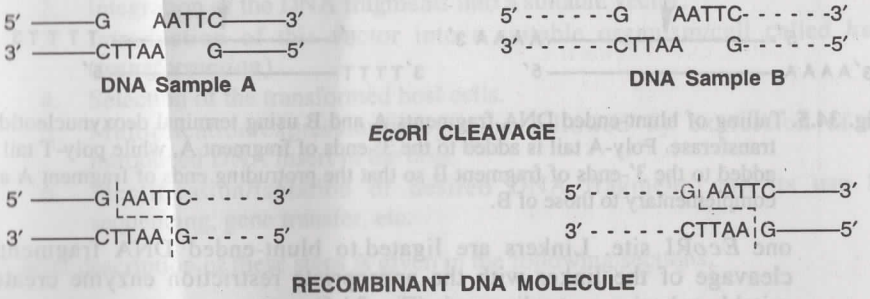


Fig. 34.4. Two distinct samples (A and B) of DNA are cleaved with the same restriction enzyme (*Eco*RI). The fragments from sample A readily join with those from sample B due to their cohesive (or sticky = complementary) protruding ends. The dotted line within the recognition sequence separates the base sequence belonging to the different fragments.

Modification of Cut Ends
- Cut ends modified in several ways.
- *Tailing*: Single-stranded tails added to 3′-ends; converts blunt ends into protruding ends.
- *Linkers*: Double-stranded oligonucleotides; used to convert blunts ends into protruding ends.
- *Adaptors*: Double-stranded oligonucleotides; used to join two DNA molecules having different sequences at their ends.

34.3. MODIFICATION OF CUT ENDS

Often the ends produced by restriction enzymes have to be modified for further manipulation of the fragments. Some of the modifications are summarised below.

1. ***Single-stranded tails*** (protruding ends) may be synthesized at the 3′- ends of blunt-ended fragments by the enzyme terminal deoxynucleotidyl transferase; this is called ***tailing***. This reaction can be used to generate protruding ends of defined sequence, *e.g.*, poly-A tails on the 3′- ends of the DNA insert and poly-T tails on the 3′-ends of the vector (Fig. 34.5); the protruding ends of the DNA insert and the vector will, therefore, base-pair under annealing conditions.
2. *Linker and/or adaptor molecules can be joined to the cut ends.* ***Linkers*** are short, chemically synthesized, double-stranded oligo-nucleotides, which contain within them one or more restriction endonuclease sites, *e.g.*, linker 5′ CCGAATTCGG (only one strand of the linker is shown here) contains

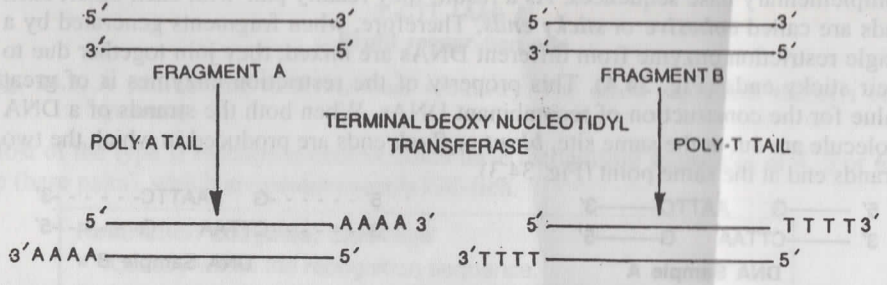


Fig. 34.5. Talling of blunt-ended DNA fragments A and B using terminal deoxynucleotidyl transferase. Poly-A tail is added to the 3′-ends of fragment A, while poly-T tail is added to the 3′-ends of fragment B so that the protruding ends of fragment A are complementary to those of B.

one *Eco*RI site. Linkers are ligated.to blunt-ended DNA fragments; cleavage of the linker with the appropriate restriction enzyme creates suitable cohesive protruding ends (Fig. 34.6).

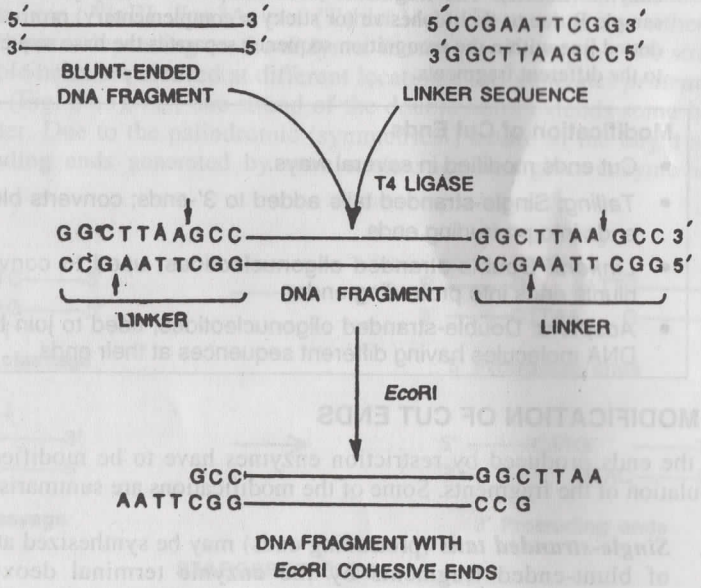


Fig. 34.6. Creation of cohesive ends on blunt-ended DNA fragments. Suitable linkers are ligated to the blunt ends by T_4 DNA ligase. The linker is then cleaved with the appropriate restriction enzyme to generate sticky ends.

3. ***Adaptors*** are short, chemically synthesized DNA double strands, which can be used to link the ends of two DNA molecules that have different sequences at their ends. They are used in conjunction with linkers or other adaptor molecules. There are different kinds of adaptors suited for different purposes.

34.4. STEPS IN GENE CLONING

The entire procedure of cloning or recombinant DNA technology may be classified into the following five steps for the convenience in description and on the basis of the chief activity performed.

1. Generation of the DNA fragments to be cloned.
2. Integration of the DNA fragments into a suitable vector.
3. Introduction of this vector into a suitable organism/cell called ***host*** (***transformation***).
4. Selection of the transformed host cells.
5. Multiplication/expression/integration followed by expression of the introduced DNA insert in the host.
6. Identification/isolation of desired DNA fragment, and its use for sequencing, gene transfer, etc.

A brief description of these steps is given in the following sections.

34.5. ISOLATION OF THE DNA FRAGMENTS

The isolation of DNA fragments to be cloned is a critical step in gene cloning. DNA inserts can be obtained from the following: (1) cDNA libraries, (2) genomic libraries, (3) chemical synthesis, and (4) amplification through polymerase chain reaction (PCR).

cDNA Library

- A collection of bacterial clones or phage lysates that has DNA sequences corresponding to mRNA isolated from a tissue/ organisms.
- *Construction of cDNA library:*
  1. mRNA is isolated.
  2. Reverse transcriptase produces DNA copy of RNA molecules; this yields DNA·RNA hybrid molecules.
  3. RNA strand is digested by RNAase H or alkaline hydrolysis.
  4. 3′-end of the DNA strand serves as template for synthesis of its complementary strand by reverse transcriptase.
  5. The hairpin is opened by an enzyme, *e.g.,* S1 nuclease.
  6. The cDNA duplexes so produced are used for cloning.
- Use of cDNA is absolutely essential when eukaryotic genes are to be expressed in prokaryotes.

34.5.1. cDNA Library

A ***cDNA library*** is a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism or tissue is represented as its *c*DNA insertion in a plasmid or a phage vector. ***cDNA*** is the copy or complementary DNA produced by using (usually) mRNA as a template. DNA copy of an RNA molecule is produced by the enzyme ***reverse transcriptase*** (RNA-dependent DNA polymerase). This enzyme performs similar reactions as DNA polymerase, and has an absolute requirement for a primer with a free 3′—OH.

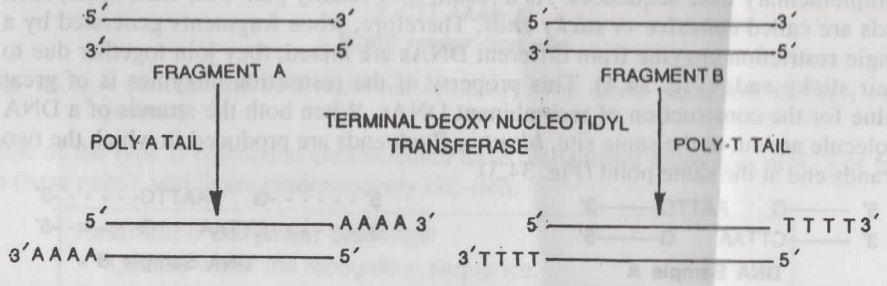


Fig. 34.5. Talling of blunt-ended DNA fragments A and B using terminal deoxynucleotidyl transferase. Poly-A tail is added to the 3′-ends of fragment A, while poly-T tail is added to the 3′-ends of fragment B so that the protruding ends of fragment A are complementary to those of B.

one *Eco*RI site. Linkers are ligated.to blunt-ended DNA fragments; cleavage of the linker with the appropriate restriction enzyme creates suitable cohesive protruding ends (Fig. 34.6).

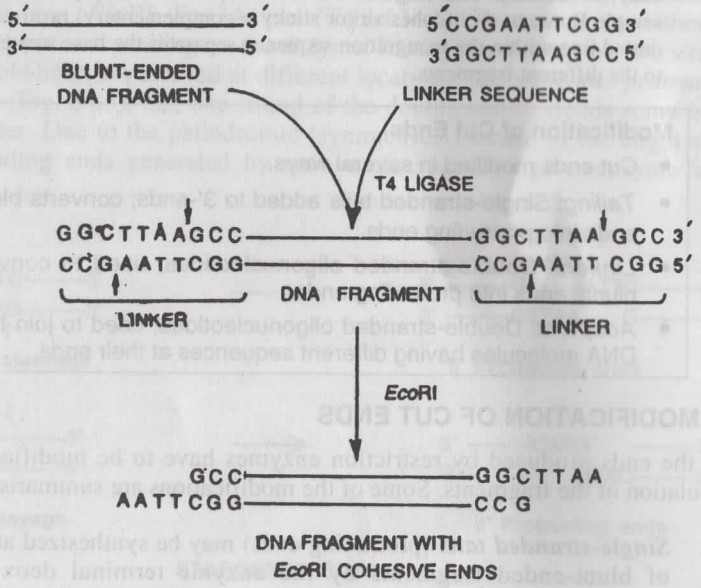


Fig. 34.6. Creation of cohesive ends on blunt-ended DNA fragments. Suitable linkers are ligated to the blunt ends by T_4 DNA ligase. The linker is then cleaved with the appropriate restriction enzyme to generate sticky ends.

3. ***Adaptors*** are short, chemically synthesized DNA double strands, which can be used to link the ends of two DNA molecules that have different sequences at their ends. They are used in conjunction with linkers or other adaptor molecules. There are different kinds of adaptors suited for different purposes.

34.4. STEPS IN GENE CLONING

The entire procedure of cloning or recombinant DNA technology may be classified into the following five steps for the convenience in description and on the basis of the chief activity performed.

1. Generation of the DNA fragments to be cloned.
2. Integration of the DNA fragments into a suitable vector.
3. Introduction of this vector into a suitable organism/cell called ***host*** (***transformation***).
4. Selection of the transformed host cells.
5. Multiplication/expression/integration followed by expression of the introduced DNA insert in the host.
6. Identification/isolation of desired DNA fragment, and its use for sequencing, gene transfer, etc.

A brief description of these steps is given in the following sections.

34.5. ISOLATION OF THE DNA FRAGMENTS

The isolation of DNA fragments to be cloned is a critical step in gene cloning. DNA inserts can be obtained from the following: (1) cDNA libraries, (2) genomic libraries, (3) chemical synthesis, and (4) amplification through polymerase chain reaction (PCR).

cDNA Library

- A collection of bacterial clones or phage lysates that has DNA sequences corresponding to mRNA isolated from a tissue/ organisms.
- *Construction of cDNA library:*
  1. mRNA is isolated.
  2. Reverse transcriptase produces DNA copy of RNA molecules; this yields DNA·RNA hybrid molecules.
  3. RNA strand is digested by RNAase H or alkaline hydrolysis.
  4. 3′-end of the DNA strand serves as template for synthesis of its complementary strand by reverse transcriptase.
  5. The hairpin is opened by an enzyme, *e.g.,* S1 nuclease.
  6. The cDNA duplexes so produced are used for cloning.
- Use of cDNA is absolutely essential when eukaryotic genes are to be expressed in prokaryotes.

34.5.1. cDNA Library

A ***cDNA library*** is a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism or tissue is represented as its *c*DNA insertion in a plasmid or a phage vector. ***cDNA*** is the copy or complementary DNA produced by using (usually) mRNA as a template. DNA copy of an RNA molecule is produced by the enzyme ***reverse transcriptase*** (RNA-dependent DNA polymerase). This enzyme performs similar reactions as DNA polymerase, and has an absolute requirement for a primer with a free 3′—OH.

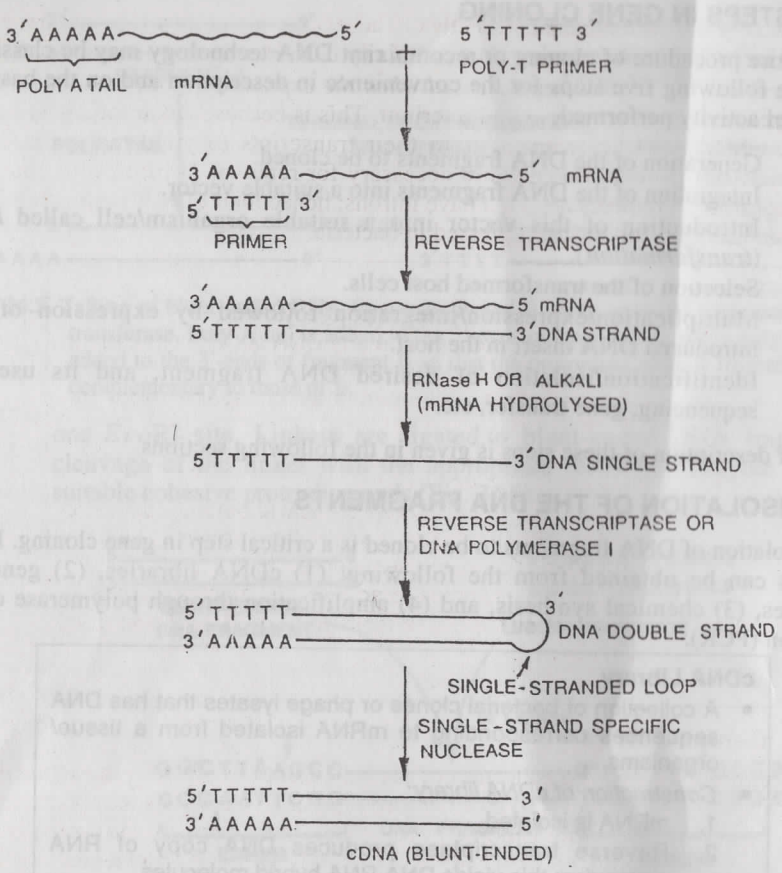


Fig. 34.7. Production of cDNA (complementary or copy DNA) from an mRNA molecule used as template by the enzyme reverse transcriptase.

When eukaryotic mRNA is used as a template, a poly-T oligodeoxynuclotide is used as the primer since these mRNAs have a poly-A tail at their 3′-ends. But special tricks are required to utilize primers for other RNAs, *e.g.*, prokaryotic mRNA, rRNA, RNA virus genomes, etc.

The appropriate oligonucleotide primer is annealed with the mRNA; this primer will base-pair to the 3′-end of mRNA (Fig. 34.7). Reverse transcriptase extends the 3′-end of the primer using mRNA molecule as a template. This produces a RNA: DNA hybrid molecule; the DNA strand of this hybrid is obviously the DNA copy (cDNA) of the mRNA strand. The RNA strand is digested either by RNAase H or alkaline hydrolysis; this frees the single stranded cDNA. Curiously, the end of this cDNA serves as its own primer and provides the free 3′—OH required for the synthesis of its complementary strand; therefore, a primer is not required for this step (Fig. 34.7). The complementary strand of cDNA single strand is synthesized by either the reverse transcriptase itself or by *E. coli* DNA polymerase I. Since the 3′-end of cDNA single strand is used as the primer for this reaction, a short hairpin loop is generated at this end. The hairpin loop is cleaved by a single strand-specific

nuclease to yield a regular DNA duplex. The cDNA molecules thus obtained are inserted into a suitable vector, and are ordinarily cloned in a suitable strain of *E. coli*.

Use of cDNA is absolutely essential when the expression of an eukaryotic gene is required in a prokaryote, e.g., a bacterium. This is because eukaryotic genes have introns, which must be removed from their transcripts to yield mature mRNA. Bacteria do not possess the enzymes necessary for removal of introns. In contrast, functional mRNA molecules do not have introns; hence their cDNA is also free from introns and can be cloned and expressed in bacteria.

Genomic Library
- A collection of bacterial clones or phage lysates.
- Ideally, has at least one copy of every DNA segment present in the genome of the concerned organism.
- Some DNA segments, however, may be absent from the library.
- *Construction*:
  1. Total genomic DNA extracted.
  2. DNA cut into fragments by mechanical shearing, sonication or partical digestion with a restriction enzyme (usually having 4 base recognition sequence).
  3. DNA fragments subjected to agarose gel electrophoresis.
  4. Fragments of suitable size isolated from the gel and used for cloning.
- This is the *shotgun approach.*

34.5.2. Genomic Library

A ***genomic library*** is a collection of plasmid clones or phage lysates containing recombinant DNA molecules so that the sum total of DNA inserts in this collection, ideally, represents the entire genome of the concerned organism. However, certain DNA fragments should be expected to be under- or over-represented, or even missing, from genomic libraries.

For preparation of a genomic library, the total genomic DNA of an organism is extracted. The DNA is broken into fragments of appropriate size either by mechanical shearing, sonication, or by using a suitable restriction endonuclease for partial digestion of the DNA. Generally, restriction enzymes having 4-base (tetrameric) recognition sequences are employed, since fragments produced in partial digests with such enzymes are more likely to be of appropriate size for cloning. Single or mixed digestions with the enzymes *Alu*I (AG/CT), *Hae*III (GG/CC) or *Sau*3A (/GATC) have been used for constructing genomic libraries.

The partial digests of genomic DNA are subjected to agarose gel electrophoresis or sucrose gradient centrifugation for separation from the mixture of fragments of appropriate size. These fragments are then inserted into a suitable vector for cloning. This constitutes the ***shotgun approach*** to gene cloning. In principle, any vector can be used, but λ vectors and cosmids have been the most commonly used. The vectors containing the inserts are cloned in a suitable bacterial host.

34.5.3. Identification of the Desired Clone

Identification of bacterial colony containing the desired DNA fragments from among those making up the library generally employs a suitable hybridization probe

in colony hybridization. The probe may be mRNA of the gene, cDNA of its mRNA, homologous gene from another organism, or a synthetic oligonucleotide representing the sequence of a part of the desired gene/DNA fragment. In order to detect hybridization, the probe must be labelled, usually with a radioactive isotope.

Chemical Synthesis of DNA

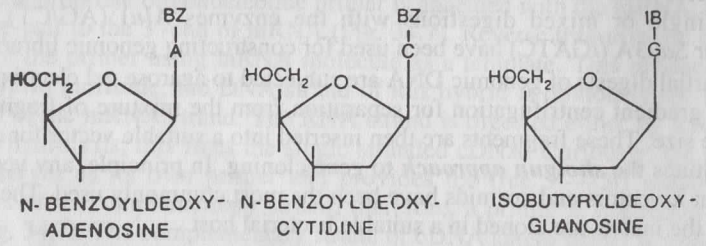
- DNA synthesized through chemical reactions.
- Uses deoxyribosides as the raw material.
- Three approaches:
  1. *Phosphodiester approach*: First to be successful; used by Khorana and others.
  2. *Phosphotriester approach*: More convenient and successful.
  3. *Phosphitetriester approach*: Most successful approach.

34.6. CHEMICAL SYNTHESIS OF GENE

If the amino acid sequence of the protein (or the base sequence of mRNA) produced by a gene is known; it enables the deduction of base sequence of the concerned gene on the basis of the codons for the various amino acids. Once the base sequence of a gene is deduced, a polynucleotide of the same base sequence can be synthesized either chemically, or even enzymatically.

There are three distinct methods for chemical synthesis of DNA: (1) phosphodiester approach, (2) phosphotriester or phosphate triester approach, and (3) phosphite triester or phosphoramidite approach. The ***phosphodiester approach*** was the first successful strategy used by Khorana and coworkers for synthesis is of alanine tRNA in 1970. It was soon replaced by the ***phosphotriester approach***, which was more successful than the phosphodiester approach; the letter was automated to give rise to the first gene machines. ***Phosphite triester method*** is the recent most strategy of DNA synthesis. It is also the most successful approach, and is briefly described below.

In the phosphite triester methods, the amino groups of the nucleosides deoxyadenosine and deoxcytidine are usually benzoylated, and that of deoxyguanosine is protected by an isobutyryl group (Fig. 34.8); thymidine requires no protection. The 5′—OH group is protected by dimethoxytrityl (Fig. 34.9), commonly abbreviated as DMTr or $(MeO)_2Tr$ (Table 34.2). The amino groups of the



N-BENZOYLDEOXY-ADENOSINE N-BENZOYLDEOXY-CYTIDINE ISOBUTYRYLDEOXY-GUANOSINE

Fig. 34.8. Protection of the amino groups of nucleosides (thymine needs no protection). Bz = Benzoyl group, IB = isobutyryl group (attached to the free amino groups of DNA bases).

bases are freed by mild alkaline hydrolysis, while $(MeO)_2Tr$ is removed by gentle acid hydrolysis.

TABLE 34.2
Agents used for the protection and deprotection of different reactive groups of nucleosides used in chemical synthesis of DNA

Protected group	*Protected by*	*Protecting group removed by*
The free $—NH_2$ group of		
deoxyadenosine	Benzoate	Mild alkaline hydrolysis
deoxycytidine	Benzoate	Mild alkaline hydrolysis
deoxyguanosine	Isobutyrate	Mild alkakine hydrolysis
thymidine	Not required	Not required
Hydroxyl groups of deoxyribose: 5′OH	Dimethoxytrityl [$(MeO)_2$ Tr] or DMTr	Gentle acid hydrolysis by benzene sulfonic acid or $ZnBr_2$

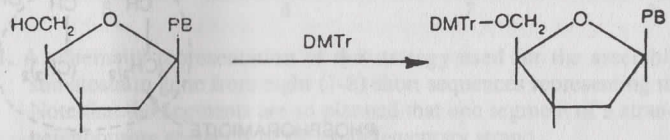


Fig. 34.9. Protection of the 5′ —OH by addition of dimethoxytrityl (DMTr). PB = amino group protected base (see Fig. 34.8).

Phosphite Triester Approach

- Amino groups of A and C protected by benzoate; of G by isobutyrate; T needs no protection.
- 5′ —OH protected by DMTr.
- *Procedure*:
  1. Free 3′ —OH of first nucleoside is linked to a solid support, *e.g.*, silica gal.
  2. 5′—OH of this nucleoside is freed by gentle acid hydrolysis, which removes DMTr.
     (*Subsequent nucleosides are used as 3′-phosphoramidites.*)
  3. The second nucleoside 3′-phosphoramidite is added and is linked to the first one by adding tetrazole.
     (The above reaction generates a phosphite linking the two nucleosides.)
  4. Phosphite is oxidized to phosphate by iodine.
  5. Reactions 2-4 are repeated to yield 50 bases long oligonucleotides.
  6. Finally, the protective groups are removed.

Nucleosides having protected bases and 5′ —OH groups are the basic materials. The 3′ —OH of the first nucleoside is fixed to a solid support and, then, its 5′ —OH is freed by gentle acid hydrolysis. The subsequent nucleosides are used as

3′-phosphoramidities, which are readily produced by coupling the nucleosides with di-isopropyl ammonium tetrazolide. The phosphoramidites are stable and efficient coupling agents, and are readily synthesized.

The desired nucleoside 3-phosphoramidite is now added and is activated for coupling by the addition of tetrazole; the immediate product is a phosphite, which is oxidized to phosphate by iodine (I_2). This phosphate remains as a triester (Fig. 34.10). The 5′ —OH of the second nucleotide is now freed, and the third desired nucleoside 3′-phosphoramidite is added and joined to the growing chain. In this manner, the oligonucleotide chain is elongated. Finally, the various protective groups are removed and the chain is freed from the solid support by alkaline hydrolysis.

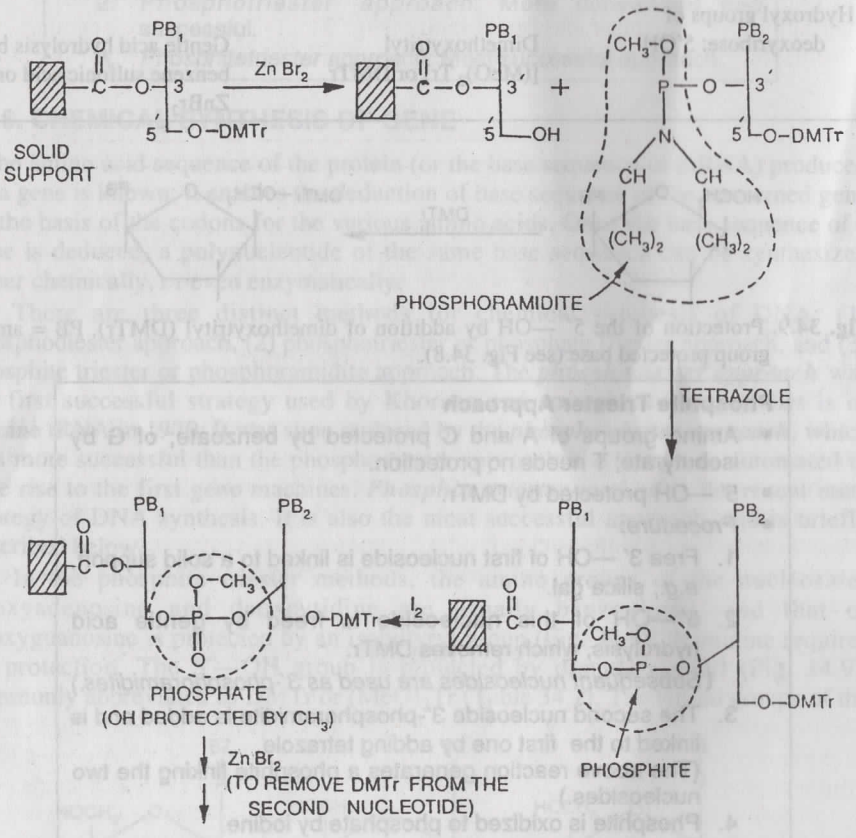


Fig 34.10. The phosphite triester approach to oligonucleotide synthesis. Phosphoramidite may be depicted as $CH_3—O—\underset{\displaystyle R}{\underset{|}{P}}—O—$ for simplicity. DMTr = dimethoxytrityl,

PB_1, PB_2, etc. = protected bases of first, second, etc. of the nucleotides added to the elongating chain.

This approach, using silica beads or controlled pore glass beads as solid support, is used for automated synthesis of oligonucleotides. It takes less than 15 minutes for adding one nucleotide to the chain, and chains as long as 50 nucleotides can be prepared in good yields. The automated DNA synthesizers are popularly called ***gene machines***; they are microprocessor controlled and carry out all the operations automatically.

It is not possible to synthesize the entire gene directly, but oligonucleotides of upto about 50 nucleotides are rapidly synthesized. One of the strategies for synthesizing the complete gene is to synthesize short stretches of the two DNA strands in such a way that each segment of one strand overlaps two segments of the other strand (Fig. 34.11). Therefore, the segments of the two strands become properly aligned due to complementary base-pairing among the various segments of the two strands. For example, the somatostatin gene was synthesized as eight overlapping segments of the two strands (Fig. 34.11); since these segments had 5′—OH, they were phosphorylated by polynucleotide kinase prior to their ligation.

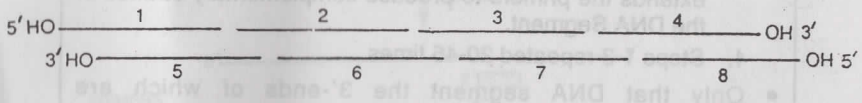


Fig. 34.11. A schematic representation of the strategy used for the assembly of complete somatostatin gene from eight (1-8) short sequences representing the two strands. Note that the segments are so planned that one segment of a strand overlaps two neighbouring segments of the complementary strand.

34.6.1. Applications of Synthetic Oligonucleotides

Synthetic oligonuclotides are used for the following purposes: (1) sequences of 12-20 bases are used as hybridization probes when other probes are not available. (2) They are used as primers for (*a*) cDNA preparation using reverse transcriptase, (*b*) amplification of DNA segments using PCR and (*c*) enzymatic DNA sequencing. (3) Oligonucleotides are also used as linkers and adaptors in gene cloning (Section 34.3). Finally, (4) they are used to produce complete gene sequences by linking them in a defined order. Efforts are being made (5) to use antisense oligonuclotides for the treatment of human diseases.

34.7. GENE AMPLIFICATION THROUGH POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) technique, developed by Kary Mullis in 1985, is extremely powerful. It generates microgram (µg) quantities of DNA copies (upto billion copies) of the desired DNA (or RNA) segment, present even as a single copy in the initial preparation, in a matter of few hours. The PCR process has been completely automated and compact thermal cyclers are available in the market.

The PCR is carried out *in vitro*. It utilizes (1) a DNA preparation containing the desired segment to be amplified, (2) two nucleotide primers (about 20 bases long) specific, *i.e.*, complementary, to the two 3′ borders (the sequences present at the 3′-ends) of the desired segment, (3) the four deoxynucleoside triphosphates, *viz*, TTP (thymidine triphosphate), dCTP (deoxycyctidine triphosphate), dATP (deoxy-adenosine triphosphate) and dGTP (deoxyguanosine triphosphate), and (*a*) a heat

Polymerase Chain Reaction

- Developed by K. Mullis in 1985.
- *Reaction mixture*:
  1. DNA preparation, a segment from which is to be amplified.
  2. Two nucleotide primers specific to the 3′-ends of the concerned DNA segment.
  3. TTP, dCTP, dGTP and dATP.
  4. A heat stable DNA polymerase, *e.g., Taq, Pfu,* or *Vent*
- *Procedure*:
  1. *Step*1: Reaction mixture heated to 90-98°C (*denaturation*).
  2. *Step* 2: Mixture cooled to 40-60°C; primers anneals to complementary sequences in the DNA (*annealing*).
  3. *Step* 3: Temperature raised to ~ 70°C; DNA polymerase extends the primers to produce complementary strands of the DNA Segment.
  4. Steps 1-3 repeated 20-45 times.
- Only that DNA segment the 3′-ends of which are complementary to the two primers used will be amplified.

stable DNA polymerase, *e.g., Taq* (isolated from the bacterium *Thermus acquaticus*), *Pfu* (from *Pyrococcus furiosus*) and *Vent* (from *Thermococcus litoralis*) polymerases. *Pfu* and *Vent* polymerases are more efficient than the *Taq* polymerase.

34.7.1. Procedure of PCR

At the start of PCR, the DNA from which a segment is to be amplified, an excess of the two primer molecules, the four deoxyriboside triphosphates and the DNA polymerase are mixed together in the reaction mixture. The following operations are now performed sequentially (Fig. 34.12).

Step 1. The reaction mixture is heated to a temperature (usually 90-98°C) that assures DNA denaturation. This is the ***denaturation step.*** This step is of 4 min in the first cycle, but of only 2 min in the subsequent cycles.

Step 2. The mixture is now cooled to a temperature (generally 40-60°C) that permits annealing of the primer to the complementary sequences in the DNA; these sequences are located at the 3′-ends of the two strands of the desired segment. This step is called ***annealing*** (Fig. 34.12). The annealing temperature depends on the length and the G.C content of the primer being used.

Step 3. The temperature is now so adjusted that the DNA polymerase synthesizes the complementary strands by utilizing 3′—OH of the primers; this reaction is the same as that occurs *in vivo* during replication of the leading strand of a DNA duplex. The primers are extended towards each other so that the DNA segment lying between the two primers is copied; *this is ensured by employing primers complementary to the 3′-ends of the segment to be amplified.*

In case of *Taq* polymerase, the optimum temperature for synthesis is between 70° and 75°C; the temperature of reaction mixture is, therefore, adjusted to this temperature. This situation has the following additional advantages. *Between 70-*

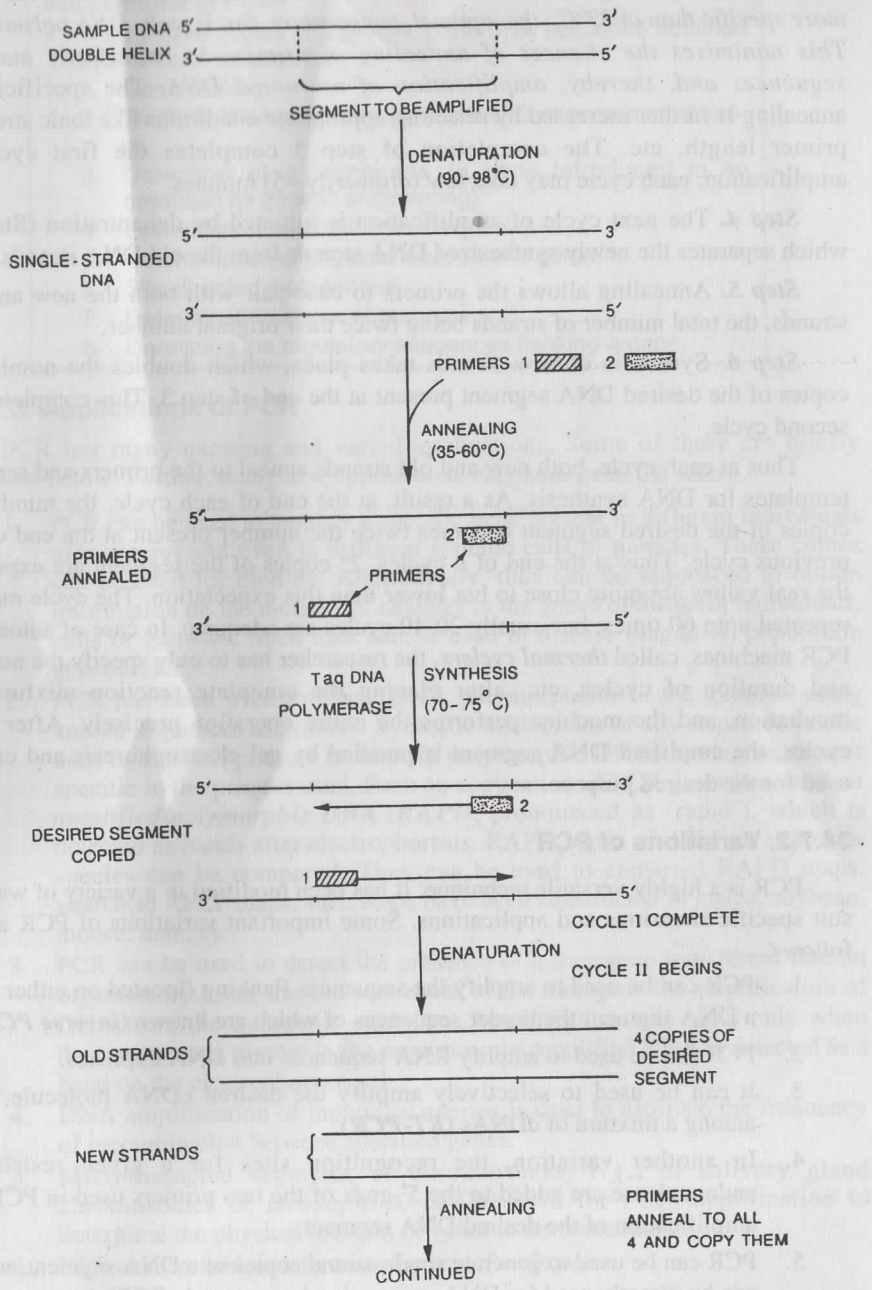


Fig. 34.12. A schematic representation of the three steps performed during PCR and their consequences. Note that the two primers used are complementary to the 3′-end sequences of the DNA segment to be amplified.

75°C, the base pairing between about 20 bases long primers and the DNA is much more specific than at 37°C, the optimal temperature for E. coli *DNA polymerase. This minimises the chances of annealing of primers to imperfectly matched sequences and, thereby, amplification of unwanted DNA.* The specificity of annealing is further increased by selecting appropriate conditions like ionic strength, primer length, etc. The completion of step 3 completes the first cycle of amplification; each cycle may take few (ordinarily ~5) minutes.

Step 4. The next cycle of amplification is initiated by denaturation (Step 1), which separates the newly synthesized DNA strands from the old DNA strands.

Step 5. Annealing allows the primers to base pair with both the new and old strands, the total number of strands being twice their original number.

Step 6. Synthesis of new strands takes place, which doubles the number of copies of the desired DNA segment present at the end of step 3. This completes the second cycle.

Thus at each cycle, both new and old strands anneal to the primers and serve as templates for DNA synthesis. As a result, at the end of each cycle, the number of copies of the desired segment becomes twice the number present at the end of the previous cycle. Thus at the end of n cycles, 2^n copies of the segment are expected; the real values are quite close to but lower than this expectation. The cycle may be repeated upto 60 times, but usually 20-40 cycles are adequate. In case of automated PCR machines, called ***thermal cyclers***, the researcher has to only specify the number and duration of cycles, etc. after placing the complete reaction mixture for incubation, and the machine performs the entire operation precisely. After PCR cycles, the amplified DNA segment is purified by gel electrophoresis and can be used for the desired purpose.

34.7.2. Variations of PCR

PCR is a highly versatile technique; it has been modified in a variety of ways to suit specific situations and applications. Some important variations of PCR are as follows.

1. PCR can be used to amplify the sequences flanking (located on either side) a DNA segment the border sequences of which are known (***inverse PCR***).
2. PCR can be used to amplify RNA sequences into DNA duplexes.
3. It can be used to selectively amplify the desired cDNA molecule from among a mixture of cDNAs (***RT-PCR***).
4. In another variation, the recognition sites for a given restriction endonuclease are added to the 5′-ends of the two primers used in PCR for amplification of the desired DNA segment.
5. PCR can be used to generate single-strand copies of a DNA segment, which can be directly used for DNA sequencing (***asymmetric PCR***).
6. PCR can be used to introduce desired mutations at specified sites in a gene (***site-directed mutagenesis***).

Applications of PCR

- The multiple copies of a specific gene/DNA sequence obtained by PCR can be used for:
  1. Gene cloning.
  2. To study DNA polymorphism as RAPD or a variation thereof.
  3. Detection of a *transgene* (a gene introduced in an organism by genetic engineering).
  4. Estimation of recombination frequency.
  5. Determination of physical location of a gene.
  6. Site-directed mutagenesis.
  7. Determination of sex of embryos.
  8. Dermining the regulatory sequences flanking a gene.

34.7.3. Applications of PCR

PCR has many exciting and varied applications. Some of these are briefly outlined below. Further, many new applications may emerge in the future.

1. PCR can be used to amplify a specific gene present in different individuals of a species and even in different somatic cells or gametes. These copies can be used for cloning. Alternatively, they can be sequenced to obtain information on the mutational changes in the genes of different individuals, cells or gametes. Such data can be used in disease diagnosis, population genetics, etc.
2. PCR has been used to study DNA polymorphism in the genome using known sequences as primers. Synthetic nucleotides of any sequence can be used as random primers to amplify polymorphic DNAs having sequences specific to the primers used. Such an application of PCR generates ***random amplified polymorphic DNA*** (***RAPD***, pronounced as 'rapid'), which is detected as bands after electrophoresis. RAPD bands of different strains or species can be compared. They can be used to construct RAPD maps, similar to RFLP maps; such maps have been constructed in maize, soybean, mouse, man, etc.
3. PCR can be used to detect the presence of a transgene transferred into an organism by using the end sequences of the transgene for amplification of DNA from the transgenic organism. Amplification will occur only when the transgene is present in the organism; the amplified DNA is detected as a band on the electrophoretic gel.
4. DNA amplification of individual sperms is used to estimate the frequency of recombination between specified genes.
5. Microdissected segments of chromosomes, *e.g.*, of salivary gland chromosomes of *Drosophila,* can be used for PCR amplification to determine the physical location of genes in chromosomes.
6. PCR can be used to determine the sex of embryos.

34.7.4. PCR Versus Gene Cloning

The PCR is a revolutionary technology. (1) It is more efficient (as it needs much less amount of the desired DNA; a single copy is enough), (2) does not require

75°C, the base pairing between about 20 bases long primers and the DNA is much more specific than at 37°C, the optimal temperature for E. coli *DNA polymerase. This minimises the chances of annealing of primers to imperfectly matched sequences and, thereby, amplification of unwanted DNA.* The specificity of annealing is further increased by selecting appropriate conditions like ionic strength, primer length, etc. The completion of step 3 completes the first cycle of amplification; each cycle may take few (ordinarily ~5) minutes.

Step 4. The next cycle of amplification is initiated by denaturation (Step 1), which separates the newly synthesized DNA strands from the old DNA strands.

Step 5. Annealing allows the primers to base pair with both the new and old strands, the total number of strands being twice their original number.

Step 6. Synthesis of new strands takes place, which doubles the number of copies of the desired DNA segment present at the end of step 3. This completes the second cycle.

Thus at each cycle, both new and old strands anneal to the primers and serve as templates for DNA synthesis. As a result, at the end of each cycle, the number of copies of the desired segment becomes twice the number present at the end of the previous cycle. Thus at the end of n cycles, 2^n copies of the segment are expected; the real values are quite close to but lower than this expectation. The cycle may be repeated upto 60 times, but usually 20-40 cycles are adequate. In case of automated PCR machines, called ***thermal cyclers***, the researcher has to only specify the number and duration of cycles, etc. after placing the complete reaction mixture for incubation, and the machine performs the entire operation precisely. After PCR cycles, the amplified DNA segment is purified by gel electrophoresis and can be used for the desired purpose.

34.7.2. Variations of PCR

PCR is a highly versatile technique; it has been modified in a variety of ways to suit specific situations and applications. Some important variations of PCR are as follows.

1. PCR can be used to amplify the sequences flanking (located on either side) a DNA segment the border sequences of which are known (***inverse PCR***).
2. PCR can be used to amplify RNA sequences into DNA duplexes.
3. It can be used to selectively amplify the desired cDNA molecule from among a mixture of cDNAs (***RT-PCR***).
4. In another variation, the recognition sites for a given restriction endonuclease are added to the 5′-ends of the two primers used in PCR for amplification of the desired DNA segment.
5. PCR can be used to generate single-strand copies of a DNA segment, which can be directly used for DNA sequencing (***asymmetric PCR***).
6. PCR can be used to introduce desired mutations at specified sites in a gene (***site-directed mutagenesis***).

Applications of PCR

- The multiple copies of a specific gene/DNA sequence obtained by PCR can be used for:
  1. Gene cloning.
  2. To study DNA polymorphism as RAPD or a variation thereof.
  3. Detection of a *transgene* (a gene introduced in an organism by genetic engineering).
  4. Estimation of recombination frequency.
  5. Determination of physical location of a gene.
  6. Site-directed mutagenesis.
  7. Determination of sex of embryos.
  8. Dermining the regulatory sequences flanking a gene.

34.7.3. Applications of PCR

PCR has many exciting and varied applications. Some of these are briefly outlined below. Further, many new applications may emerge in the future.

1. PCR can be used to amplify a specific gene present in different individuals of a species and even in different somatic cells or gametes. These copies can be used for cloning. Alternatively, they can be sequenced to obtain information on the mutational changes in the genes of different individuals, cells or gametes. Such data can be used in disease diagnosis, population genetics, etc.
2. PCR has been used to study DNA polymorphism in the genome using known sequences as primers. Synthetic nucleotides of any sequence can be used as random primers to amplify polymorphic DNAs having sequences specific to the primers used. Such an application of PCR generates ***random amplified polymorphic DNA*** (***RAPD***, pronounced as 'rapid'), which is detected as bands after electrophoresis. RAPD bands of different strains or species can be compared. They can be used to construct RAPD maps, similar to RFLP maps; such maps have been constructed in maize, soybean, mouse, man, etc.
3. PCR can be used to detect the presence of a transgene transferred into an organism by using the end sequences of the transgene for amplification of DNA from the transgenic organism. Amplification will occur only when the transgene is present in the organism; the amplified DNA is detected as a band on the electrophoretic gel.
4. DNA amplification of individual sperms is used to estimate the frequency of recombination between specified genes.
5. Microdissected segments of chromosomes, *e.g.*, of salivary gland chromosomes of *Drosophila,* can be used for PCR amplification to determine the physical location of genes in chromosomes.
6. PCR can be used to determine the sex of embryos.

34.7.4. PCR Versus Gene Cloning

The PCR is a revolutionary technology. (1) It is more efficient (as it needs much less amount of the desired DNA; a single copy is enough), (2) does not require

difficult to store and costly restriction enzymes, ligase, and vector DNA, which reduces the cost drastically; (3) needs far less work, time and skill; (4) and has many more applications than gene cloning. Typically, gene cloning experiments take 2–4 days, while PCR takes upto 4 hr. In addition, (5) PCR is fully automated, while gene cloning is not. But for PCR one does need a thermal cycler (cost around Rs. 2 lac) or PCR machine. It is expected that PCR will eventually take over most of the applications of gene cloning and will find many more applications as well. But gene cloning is likely to remain a basic technique in recombinant DNA technology because the first copies of a gene are ordinarily obtained by gene cloning only.

QUESTIONS

1. Define gene cloning. List the various approaches for obtaining the DNA segments for cloning and describe any one of them in detail.
2. Briefly describe the procedure for construction of cDNA and genomic libraries and explain their applications.
3. What are the various approaches for chemical synthesis of DNA. Describe the most successful approach in some detail. Outline a strategy for synthesis of a complete gene, and list the various uses of synthetic oligonucleotides.
4. Briefly describe the various steps of PCR, and list its variations. Enumerate the various applications of PCR, and compare it with gene cloning.
5. What are restriction enzymes ? Explain their properties relevant to and applications for gene cloning.
6. Write Short notes on:
   (*i*) restriction enzymes, (*ii*) cohesive and blunt ends
   (*iii*) linkers and adaptors, (*iv*) cDNA library,
   (*v*) PCR, (*vi*) genomic library,
   (*vii*) recognition sequence, (*viii*) recombinant DNA technology.

35

Genetic Transformation of Prokaryotes

35.1. INTRODUCTION

DNA segments/genes (***DNA inserts***) isolated from any organism are first cloned , ordinarily, in *E. coli.* For this purpose, the DNA segments are first integrated into a suitable vector, which is then introduced into *E. coli* cells. *E. coli* cells containing the vector (into which the DNA inserts have been integrated) are then selected using appropriate techniques. These operations are briefly described in this chapter.

35.2. VECTORS

A Vector is a DNA molecule that has the ability to replicate in an appropriate host cell, and into which the DNA fragment to be cloned (called ***DNA insert***) is integrated for cloning. Therefore, a vector must have an origin of DNA replication (denoted as *ori*) that functions efficiently in the concerned host cell.

A Good Vector
- Replication autonomous.
- Easy purification.
- Easy transformation.
- Easy detection / selection (due to suitable marker genes).
- Easy selection/identification of cells having recombinant DNA.
- Unique restriction sites.
- In case of expression vectors, suitable regulatory sequences.

NATURALLY OCCURRING VECTORS ARE POOR CLONING VECTORS.

35.2.1. Properties of A Good Vector

A good vector must have the following properties.

1. It should be able to replicate autonomously, *i.e.*, independent of the host chromosome replication.
2. It should be easy to isolate and purify.
3. It should be easily introduced into the host cells.
4. The vector should have suitable marker genes that allow easy detection and/ or selection of the transformed host cells.
5. The cells transformed with the vector molecules containing the DNA insert

(***recombinant DNA***) should be identifiable or selectable from those transformed by the vector molecules only.

6. A vector should contain unique target sites for as many restriction enzymes as possible; this is necessary for gene cloning.
7. When expression of the DNA insert is desired, the vector should contain suitable control elements, like promoter, operator and ribosome binding sites, etc. Such vectors are called ***expression vectors.***

Therefore, the choice of vector depends largely on the host species into which the DNA insert or gene is to be cloned. Most naturally occurring DNA molecules do not have all the required functions for a vector. Threfore, useful vectors have been created by joining together segments performing specific functions from two or more natural entities. A brief description of some of the important vectors used in *E. coli* is given below.

Vector Types

- *Cloning vectors*: All vectors used for gene cloning.
- *Expression vectors*: Used for expression of, *i.e.*, protein production by, the DNA insert; contain all the regulatory sequences needed for gene expression in the host concerned. Two types:
  1. *Translational fusion vectors*: DNA insert encoded protein is fused with a vector encoded polypeptide.
  2. *Transcriptional fusion vectors*: DNA insert encoded protein is free from any other sequence; only vector encoded RNA is fused with the DNA insert encoded RNA.

35.2.2. Cloning And Expression Vectors

All vectors used for propagation of DNA inserts in a suitable host are called ***cloning vectors.*** But when a vector is designed for the expression of, *i.e.*, production of the protein specified by, the DNA insert is known as *expression vector.* As a rule, such vectors contain at least the regulatory sequences, *i.e.*, promoters, operators, ribosomal binding sites, etc., having optimum function in the chosen host. It is desirable that all cloning vectors have relaxed replication control so that they can produce multiple copies per host cell.

When an eukaryotic gene is to be expressed in a prokaryote, the eukaryotic coding sequence has to be placed after prokaryotic promoter and ribosome binding site since the regulatory sequences of eukaryotes are not recognised in prokaryotes. In addition, eukaryotic genes, as a rule, contain introns (noncoding regions) present within their coding regions. These introns must be removed to enable the proper expression of eukaryotic genes since prokaryotes lack the machinery needed for intron removal from the RNA transcripts. When eukaryotic genes are isolated as cDNA, they are intron-free and, hence, suitable for expression in prokaryotes.

Several strategies have been attempted for the construction of expression vectors using regulatory sequences of the appropriate hosts. These approaches may be grouped into the following two broad categories.

1. Construction of vectors allowing the synthesis of fusion proteins comprising amino acids encoded by a sequence in the vector and those encoded by the DNA insert (***translational fusion***).

2. Development of vectors permitting the synthesis of pure proteins encoded exclusively by the DNA inserts (***transcriptional fusion***).

Examples of the first strategy producing fusion proteins are the expression of rat insulin, rat growth hormone, structural protein VP1 of foot and mouth disease virus, human growth hormone, etc. Some examples of the second approach producing unique proteins are rabbit β- globin, small *t*-antigen of SV40, human fibroblast interferon, human IGF-I protein, etc. It may be pointed out that the undesired amino acids encoded by the vector sequence in cases of translational fusion must be removed from the fusion proteins by a suitable chemical cleavage.

Several other problems are faced when eukaryotic genes are expressed in a prokaryotic system, *e.g.*, removal of signal sequences from precursor proteins to obtain active mature protein molecules, etc. Various strategies are being rapidly devised to effectively overcome these problems.

***E. coli* Vectors**

- Constructed from plasmids or phages.
- *Types of Vectors* :
  1. plasmids,
  2. bacteriophages,
  3. cosmids,
  4. phasmids,
  5. bacterial artificial chromosomes (BACs),
  6. shuttle vectors.
- Used for gene cloning in *E. coli.*
  ALL GENES ARE ORDINARILY FIRST CLONED IN *E. COLI.*

35.3. *E. coli* VECTORS

Bacteria are the hosts of choice for DNA cloning. Among them, *E. coli* occupies a prominent position since cloning and isolating DNA inserts for structural analysis is the easiest in this host. Therefore, ***the initial cloning experiments are generally carried out in* E. coli.** The *E. coli* strain K12 is the most commonly used host. *E. coli* supports several types of vectors, some natural, some constructed, which can be grouped as follows: (1) plasmids, (2) bacteriophages (both are natural entities), (3) cosmids, (4) phasmids, (5) bacterial artificial chromosomes (BACs), and (6) shuttle vectors (the last four are constructed by man).

A Good Host

- Easy to transform.
- Supports recombinant DNA replication.
- Lacks active restriction enzymes.
- Lacks methylases.
- Deficient in recombination.
- Lacks elements that interfere with the replication of recombinant DNA.

35.3.1. Properties of a Good Host

A good host should (1) be easy to transform, (2) support the replication of

recombinant DNA, (3) be free from elements that interfere with the replication of recombinant DNA, (4) lack active restriction enzymes, *e.g., E. coli* K12 substrain HB 101, (5) not have methylases, and (6) be deficient in normal recombination function. These properties ensure production of large number of copies of DNA insert in its original unmodified state.

35.3.2. Plasmid

A ***plasmid*** is a DNA molecule, other than the bacterial chromosome, that is capable of independent replication and transmission. Plasmids are circular and may exist either independent of or may become integrated into the bacterial chromosome; generally, they are not essential for the host cell except under specific environments. There are several types of bacterial plasmids, *e.g.,* F plasmids, Col plasmids, R plasmids, etc. Some plasmids are ***conjugative*** or ***transmissible***; they mediate DNA transfer through conjugation and, as a result, spread rapidly among the bacterial cells of a population, *e.g.,* F plasmids, many R plasmids and some Col plasmids. Some other plasmids are ***nonconjugative***, and they do not mediate DNA transfer through conjugation, *e.g.,* many R plasmids and most Col plasmids.

Each plasmid is maintained in the bacterial cell at a characteristic copy number mainly due to its replication control system. In this respect, the plasmids are of two types: (1) single copy and (2) multicopy plasmids. The replication control of ***single copy plasmids*** is the same as that of their bacterial host chromosomes so that they replicate and segregate with the bacterial chromosome; this is called ***stringent replication control.*** In contrast, the replication control of ***multicopy plasmids*** is different from that of their bacterial host genome so that they undergo more than one replication for each replication of their host genome; this is referred to as ***relaxed replication control.***

35.4. PLASMID VECTORS

Many different *E. coil* plasmids are used as vectors. The natural plasmids have been modified, shortened, reconstructed and recombined to create useful vectors.

TABLE 35.1
Antibiotic resistance genes found in R plasmids, their proteins and mechanism of antibiotic resistance

Antibiotic (gene conferring resistance)	*Protein produced by the gene*	*Mechanism of resistance*
Ampicillin (*amp*)	Penicillinase or β-lactamase	Hydrolysis of C-N bond in β-lactam ring.
Kanamycin (*kan*)	Kanamycin acetyltransferase*	N-acetylation of the antibiotic.
Neomycin (*neo*)	Aminogycoside phosphotransferase *	O-phosphorylation of the antibiotic.
Streptomycin (*str*)	Streptomycin phosphotransferase	Phosphorylation of the —OH on the antibiotic.
	Streptomycin adenylate synthetase	Adenylation of the —OH on the antibiotic.

* The antibiotics kanamycin and neomycin are related; hence *neo* product also inactivates kanamycin, and *kan* product inactivates neomycin as well.

An ideal plasmid vector must have the following functions: (1) a minimum amount of DNA, (2) relaxed replication control, (3) at least two selectable (Table 35.1) (or one selectable and one scorable) markers, (4) only one (unique) recognition site for at least one restriction endonuclease, and (5) for easy selection of the recombinant DNA, this unique restriction site must be located within one of the two selectable markers. Almost all such vectors have been constructed from naturally occurring DNA sequences using both classical genetic (recombination) and recombinant DNA techniques.

pBR322
- Developed by Boliver Rodriguez.
- Size 4362 bp.
- *Origin* from ColE1 plasmid.
- Two selectable markers: *tet*^{r} and *amp*^{r}.
- 12 Unique restriction sites: 2 in *tet*^{r} and 4 within *amp*^{r}.
- DNA insert integrated within *tet*^{r} (or *amp*^{r}).
- Nontransformed cells eliminated by plating on ampicilin medium (or tetracycline if DNA insert is within *amp*^{r}).
- Colonies produced on ampicillin medium replica-plated on tetracycline medium.
- Colonies that fail to develop have recombinant DNA; they are picked from the master plate.

35.4.1. *p*BR322

The name *p*BR denotes the following: *p* signifies plasmid, BR is from Boliver Rodriguez, the scientist who developed *p*BR322. *p*BR322 is 4362 bp in size, and is the most widely used plasmid. It has the *origin* from *E. coli* plasmid ColE1; it permits plasmid replication even when chromosome replication and cell division are inhibited by amino acid starvation or chloramphemicol. As a result, under such conditions, each cell accumulates several thousand copies of the plasmid so that one litre of bacterial culture easily yields a milligram of plasmid DNA. It has two selectable markers (tetracycline, *tet*^{r}, and ampicillin, *amp*^{r}, resistance genes), and unique recognition sites for 12 different restriction enzymes (two, *Pst*I and *Pvu*I are located within the *amp*^{r} gene, and 4, *e.g., Bam*HI, *Sal*I, etc., within *tet*^{r} gene) (Fig. 35.1).

The presence of restriction sites within the markers *tet*^{r} and *amp*^{r} permits an easy selection for cells transformed with the recombinant DNA. Insertion of the DNA fragment into the plasmid using restriction enzyme *Pst*I or *Pvu*I, places the DNA insert within the gene *amp*^{r} and thus makes it nonfunctional. Bacterial cells containing such a recombinant DNA will be unable to grow in the presence of ampicillin, but will grow on tetracycline.

E. coli cells tranformed with such vectors are first plated on an agar medium containing tetracycline, the antibiotic within the resistance gene for which the *DNA fragment is not inserted.* Therefore, bacterial cells having the recombinant DNA are expected to be resistant to the antibiotic. This eliminates nontransformed bacterial cells; the resulting bacterial colonies will possess either recombinant or unaltered *p*BR322. The colonies so obtained are then replica-plated on agar plates containing ampicillin; all the colonies that develop on this plate will contain the unaltered

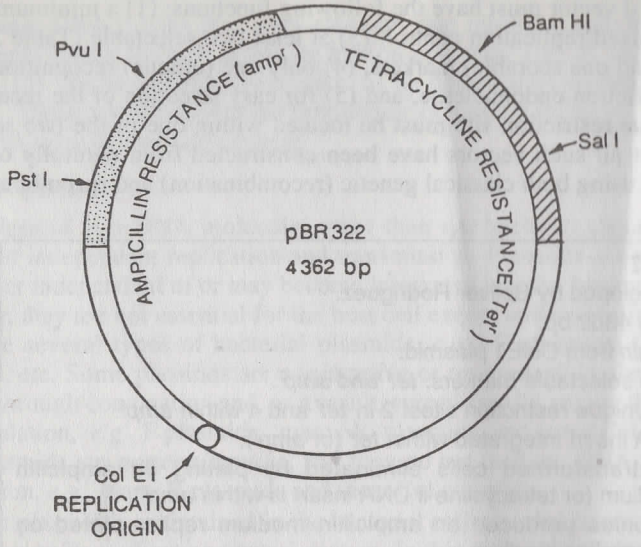


Fig. 35.1. Structural features of *p*BR322. ColE1 origin of replication allows production of multiple copies per cell, and the two resistance markers (*tet*^{r} and *amp*^{r}) allow easy selection for the recombinant DNA.

*p*BR322. Therefore, the antibiotic sensitive colonies are identified and recovered from the master plate; these colonies will have the recombinant DNA. This entire process may take 1-2 days.

Similarly, when restriction enzyme *Bam*HI or *Sal*I is used, the DNA insert is placed within gene *tet*^{r} making it nonfunctional. Bacterial cells possessing such a *p*BR322 will, therefore, grow on ampicillin but not on tetracycline (Fig. 35.2) Transformed *E. coli* cells, in this case, are first plated on agar plates containing ampicillin. The colonies so obtained are then replica-plated on agar plates having tetracycline. The colonies that fail to develop on this plate are now isolated from the master plate; these colonies contain the recombinant DNA. This feature allows an easy selection of a single bacterial cell having recombinant DNA from among 10^8 other types of cells.

35.4.2. *p*UC7

This vector is a derivative of *p*BR322 and is much smaller (2.7 kb); it has all the essential parts of *p*BR322, *e.g.*, (1) ampicillin resistance gene and (2) ColE1 origin. The second scrobale marker is due to *E. coli* gene *lacZ*α encoding the α fragment of β-galactosidase, the enzyme that hydrolyses galactose. The *E. coli* strains, *e.g.*, JM103, JM109, used as hosts for the *p*UC series vectors have the *lacZ*α sequence deleted from their *lacZ* genes. When *p*UC vector enters such an *E. coli* cell, the host genome and the plasmid encode for different parts of the β-galactosidase enzyme, which interact with each other to produce the active enzyme, enabling these cells to hydrolyse galactose. β-galactosidase also hydrolyses X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactoside) to yield a blue dye. Therefore, appropriate *lacZ*$^-$ *E. coli* cells transformed by the *p*UC vectors behave as *lacZ*$^+$ and produce blue coloured

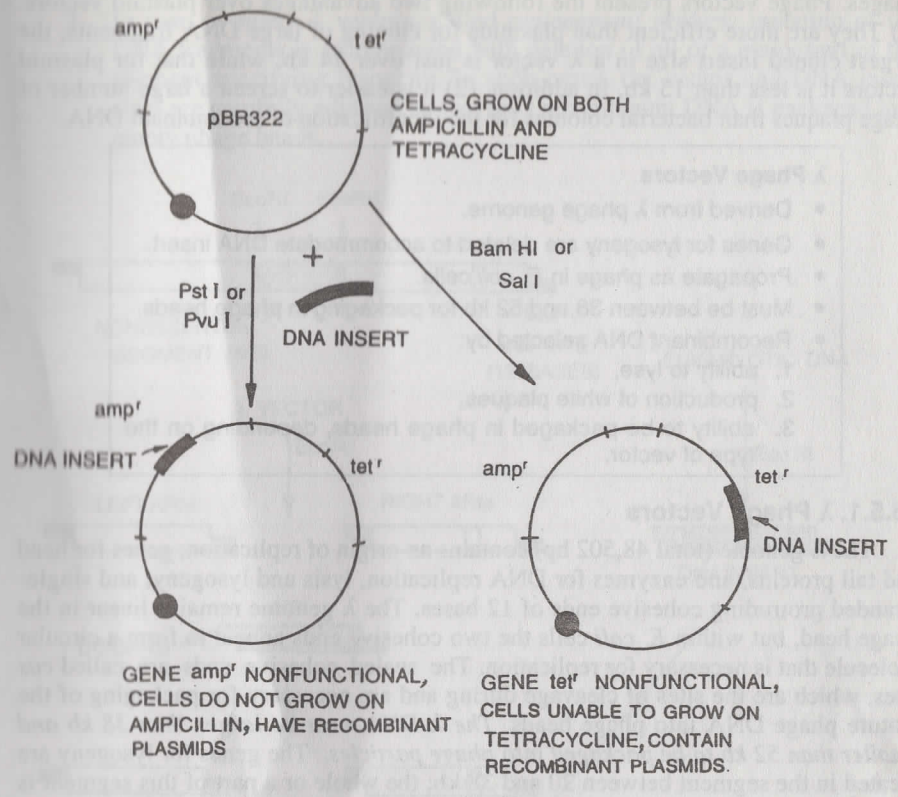


Fig. 35.2. The consequences of insertion of a DNA fragment within *amp*^{r} or *tet* gene of *p*BR322. Cells containing the recombinant DNA are resistant to only one antibiotic, they are not resistant to the antibiotic within the marker gene for which the DNA insert is integrated. In contrast, nontransformed cells are sensitive to both the antibiotics, while those transformed by the unaltered *p*BR322 are resistant to both ampicillin and tetracycline.

colonied on an X-gal containing medium. A polylinker sequence has been created (by changing the base sequence without altering the encoded amino acid sequence) within the *lacZ*α; this provides several unique restriction sites for DNA insertion. When a DNA insert is placed within the polylinker sequence, *lacZ*α expression is prevented; therefore, appropriate *E. coli* cells possessing recombinant *p*UC DNA are β-galactosidase deficient and, as a result, produce white colonies on X-gal medium. *E. coli* cells transformed with *p*UC vectors are grown on ampicillin containing medium to eliminate nontransformed cells. This medium also contains X-gal; therefore, the colonies obtained on this medium are either blue or white. The white colonies are selected as they contain the recombinant DNA.

35.5. BACTERIOPHAGE VECTORS

Bacteriophages are viruses that attack bacteria. Several becteriophages are used as cloning vectors, the most commonly used *E. coli* phages being λ (lambda) and M13

phages. Phage vectors present the following two advantages over plasmid vectors. (1) They are more efficient than plasmids for cloning of large DNA fragments; the largest cloned insert size in a λ vector is just over 24 kb, while that for plasmid vectors it is less than 15 kb. In addition, (2) it is easier to screen a large number of phage plaques than bacterial colonies for the identification of recombinant DNA.

λ Phage Vectors

- Derived from λ phage genome.
- Genes for lysogeny are deleted to accommodate DNA insert.
- Propagate as phage in *E. coli* cells.
- Must be between 38 and 52 kb for packaging in phage heads.
- Recombinant DNA selected by:
  1. ability to lyse,
  2. production of white plaques,
  3. ability to be packaged in phage heads, depending on the type of vector.

35.5.1. λ Phage Vectors

The λ genome (total 48,502 bp) contains an origin of replication; genes for head and tail proteins, and enzymes for DNA replication, lysis and lysogeny; and single-stranded protruding cohesive ends of 12 bases. The λ genome remains linear in the phage head, but within *E. coli* cells the two cohesive ends anneal to form a circular molecule that is necessary for replication. The sealed cohesive ends are called *cos* sites, which are the sites of cleavage during and are necessary for packaging of the mature phage DNA into phage heads. *The λ DNA must be larger than 38 kb and smaller than 52 kb to be packaged into phage particles.* The genes for lysogeny are located in the segment between 20 and 38 kb; the whole or a part of this segment is deleted to create λ vectors to (1) accommodate larger DNA inserts and (2) to ensure that the recombinant phage is always lytic.

These vectors have the following two basic features. (1) The vector itself can be propagated as phage in *E. coli* cells enabling preparation of vector DNA. (2) They contain restriction sites, which allow the removal of the lysogenic segment and also provide insertion site for the DNA inserts (Fig. 35.3.)

35.5.2. Selection of Recombinant λ DNA

The recombinant λ DNA has to be selected from the nonrecombinant one; this can be achieved in one of the following three ways.

1. Some λ vectors retain the lysogenic function. In such vectors, the DNA fragment is inserted within the gene *c*I (lysis repressor); in such a case, the recombinant DNA forms clear plaques, which are easily distinguishable from the cloudy plaques produced by the unaltered λ vector.
2. Some λ vectors contain the promoter, operator and *lacZ* gene of *E. coli lac* operon. Insertion of DNA fragments in such vectors is accompanied with the deletion of a major part of the *lacZ* gene. When such a vector is tested on a *lacZ*$^-$ *E. coli* lawn grown on a medium having X-gal, then unaltered vector produces blue plaques, while the recombinant DNA produces colourless plaques.

3. In case of many λ vectors, called ***replacement vectors,*** insertion of the DNA fragments is accompanied with deletion of all or a major part of the segment specifying lysogeny. In such cases, the vector and DNA insert sizes are carefully adjusted so that only recombinant DNA is packaged into empty phage heads.

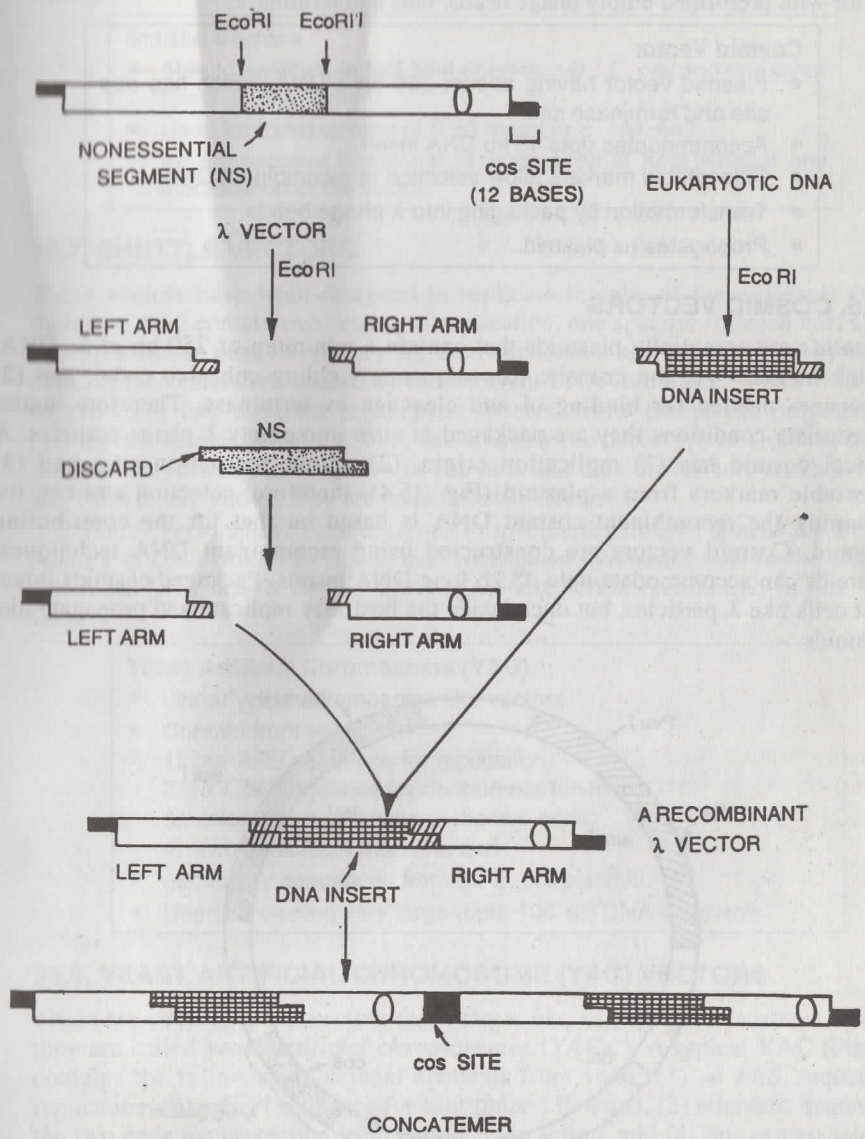


Fig. 35.3. Integration of DNA insert into a λ vector and formation of concatemer during annealing and ligation. The open circle (O) within the λ genome represents the origin of DNA replication.

35.5.3. Infection of *E. coli* cells with λ vectors

The λ vectors can be used as naked DNA to transform *E. coli* cells, or they may be packaged into empty phage heads, and used to infect the host cells; plaque formation by the latter is 100-10,000 times greater than that obtained by the former. *In vitro* packaging of λ vectors is achieved by mixing the concatameric form of vector with preformed empty phage heads, tails and terminase.

Cosmid Vector

- Plasmid vector having atleast 250 bp λ DNA, which has *cos* site and terminase site.
- Accommodates upto 45 kb DNA insert.
- Selecatable markers allow selection of recombinant DNA.
- Transformation by packaging into λ phage heads.
- Propagates as plasmid.

35.6. COSMID VECTORS

Cosmids are essentially plasmids that contain a minimum of 250 bp of λ DNA, which includes (1) the *cos* site (the sequence yielding cohesive ends), and (2) sequences needed for binding of and cleavage by terminase. Therefore, under appropriate conditions they are packaged *in vitro* into empty λ phage particles. A typical cosmid has (1) replication origin, (2) unique restriction sites and (3) selectable markers from a plasmid (Fig. 35.4); therefore, selection strategy for obtaining the recombinant cosmid DNA is based on that for the contributing plasmid. Cosmid vectors are constructed using recombinant DNA techniques. Cosmids can accommodate upto 45 kb long DNA inserts. Packaged cosmids infect host cells like λ particles, but once inside the host they replicate and propagate like plasmids.

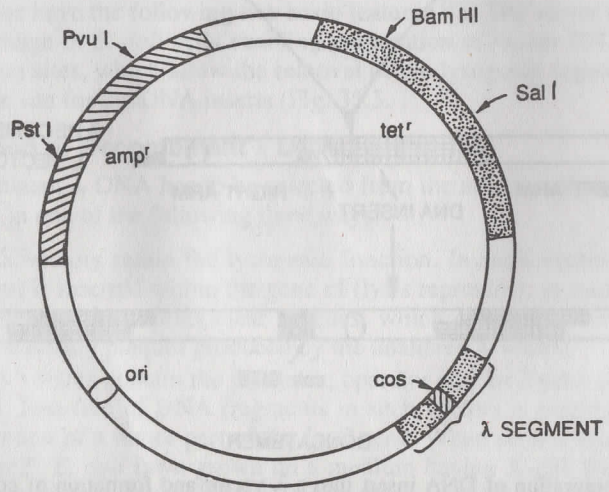


Fig. 35.4. A typical cosmid vector containing *p*BR322 modules and a λ segment having the *cos* site, and terminase binding and cleavage sequences.

The typical features of cosmids are as follows. (1) They can be used to clone DNA inserts of upto 45 kb. (2) They can be packaged into λ particles, which infect host cells; this is many-fold more efficient than plasmid transformation. (3) Selection for recombinant DNA is based on the procedure applicable to the plasmid making up the cosmid. (4) Finally, these vectors are amplified and maintained in the same manner as the contributing plasmid.

Shuttle Vectors

- Able to replicate in two host species, *viz.*, *E. coli* and one other species, *e.g.*, yeast.
- Used for initial cloning of DNA insert in *E. coli*, and
- for subsequent functional tests in another host without any modification in the vector.

35.7. SHUTTLE VECTORS

These vectors have been designed to replicate in cells of two different species; therefore, they contain two origins of replication, one specific for each host species, as well as those genes necessary for their replication and not provided by the host cells. These vectors are created by recombinant DNA techniques. Some of them can be grown in two different prokaryotic species, while others can propagate in a prokaryotic species, usually *E. coli,* and a eukaryotic one, *e.g.* yeast, plants, animals. Since these vectors can be grown in one host and then moved into another without any extra manipulation, they are called ***shuttle vectors.***

Shuttle vectors have been designed to specifically satisfy the need for the initial cloning of DNA inserts in E. coli *and subsequent functional tests in the species to which the DNA inserts belong. Most of the eukaryotic vectors are, in fact, shuttle vectors.*

Yeast Artificial Chromosome (YAC)

- Linear yeast chromosome like vectors.
- Contain from yeast:
  1. an *ARS* sequence for replication,
  2. a *CEN* sequence for contromere function,
  3. telomeric sequences at the two ends,
  4. two selectable markers, and
- necessary sequences from an *E. coli* plasmid.
- Used for cloning very large (upto 100 kb) DNA segments.

35.8. YEAST ARTIFICIAL CHROMOSOME (YAC) VECTORS

These are circular linear vectors that behave like an yeast chromosome; therefore, they are called ***yeast artificial chromosomes*** (***YACs***). A typical YAC (Fig. 35.5) contains the following functional elements from yeast: (1) an *ARS* sequence for replication, (2) a *CEN* sequence for centromeric function, (3) telomeric sequences at the two ends for protection from exonuclease action, and (4) one or two selectable marker genes, *e.g., LEU2* and *HIS3*, and (5) the necessary sequences from on *E. coli* plasmid. The ***telomeric sequence*** in yeast chromosomes is a 20-70 tandem repeat of the 6 base sequence 5'CCCCAA3' (its complementary sequence, 5'TTGGGG3'

occurs in the other strand); most likely there is a hairpin loop formation at the terminus, which makes the DNA duplex resistant to exonuclease action.

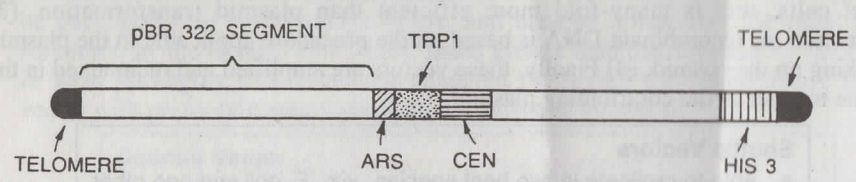


Fig. 35.5. A typical yeast artificial chromosome (YAC). Note that a YAC has telomeric sequence at its two ends and, in addition, has an *ARS* and a *CEN* sequence.

When a YAC is less than about 20 kb, the centromeric function is unable to control copy number during mitosis so that several copies of YAC accumulate per yeast cell. The centromeric function improves in YACs of 50 kb or more; YACs of 150 kb or more behave like regular yeast chromosomes. YACs are used for cloning of very large (upto 100 kb) DNA segments used for mapping of complex eukaryotic chromosomes.

35.9. VECTORS FOR ANIMALS

DNA fragments from animals are generally cloned in *E. coli* to obtain sufficient number of copies for structural analyses. However, expression of the DNA inserts has to be studied in animal cells; therefore, many of the vectors used for cloning in *E. coli* are shuttle vectors for *E. coli* and animal cells. The different vectors used in animal cells are derived from viral genomes, and are either (1) virus-like (recombinant DNA produces virions, *i.e.,* virus particles), or (2) plasmid- like vectors (these replicate but do not produce virions). (3) Some other vectors are designed to be unable to replicate. The various vectors for animals are considered in more detail in Chapter 36.

35.10. VECTORS FOR PLANTS

Plant cells do not contain any plasmid. But two plasmids, called *p*Ti and *p*Ri, that are present in the bacteria *Agrobacterium tumefaciens* and *A. rhizogenes,* respectively, provide a naturally occurring transformation system. These plasmids, in natural course, transfer a part of their DNA, called ***T-DNA,*** into the genomes of most dicot and some monocot plants. These plasmids, especially the Ti plasmid, have been used to develop a variety of vectors. In addition, genomes of many plant viruses are being developed as vectors. The purpose of plant vectors is almost always a stable transformation ordinarily in the form of integration into the plant genomes. But in the case of virus vectors, the objective is to produce large quantities of the protein encoded by the DNA insert. More information on plant vectors follows in Chapter 37.

35.11. INTEGRATION OF *DNA* INSERT INTO VECTOR

The selected vector is cut open with a restriction enzyme that has a unique (single) target site located in the sequence where the DNA insert should be integrated to facilitate the selection of recombinant DNA. There are five possible situations with

respect to the vector and the DNA insert; these situations are briefly described below.

DNA Insert Integration into Vector

1. *Both ends cohesive and compatible:* Opened up vector and DNA insert mixed under annealing conditions; vector and DNA insert ends pair; nicks ligated.
2. *Both ends cohesive and separately matched:* As in item 1.
3. *Both ends cohesive and unmatched*: The ends may be converted into matched ends by adaptors; ends may be made blunt, then subjected to tailing, ligated as such, or attached to linkers.
4. *Both ends blunt*: Vector (opened up) and DNA insert are mixed and ligated by T4 ligase; blunt ends could be converted into cohesive matched ends.
5. *One end cohesive and compatible, the other end blunt*: Opened vector and DNA insert mixed under annealing conditions: cohesive ends pair; blunt ends and nicks ligated by T4 ligase.

35.11.1. Both Ends Cohesive and Compatible

The simplest strategy concerns the presence of ***compatible cohesive ends both in the vector and the DNA insert.*** This happens when the vector is cut open by the same restriction enzyme that was used to isolate the DNA insert. The opened up vector and the DNA insert are mixed under annealing conditions, which allows pairing between the compatible cohesive ends of the vector and the DNA insert. The *nicks* (the broken covalent bonds between the 5'-phosphate and the 3'—OH of two neighbouring nucleotides within a DNA strand) remaining after annealing are sealed, joined, or ligated by DNA ligase (Fig. 35.6).

Initially, DNA ligase from *E. coli* was used, but now-a-days T4 (an *E. coli* phage) ligase is preferred. DNA inserts joined in this manner can be easily and precisely isolated from the vector using the same restriction enzyme that was used to generate the inserts and to open the vector.

In the recombinant DNA produced in this way, the DNA insert may be present in either of the two orientations relative to the sequences of the vector. If the two cohesive ends of the vector are marked as 1 and 2 and those of the DNA insert as 1' and 2', the insert may join the vector to either yield 1-1' and 2-2' junctions, or 1-2' and 2-1' junctions (Fig. 35.6). *The orientation of DNA insert within the vector is not important when only copies of the insert are to be obtained. But orientation is extremely important when expression of the insert is desired.*

35.11.2. Both Ends Cohesive and Separately Matched

One protruding end of the vector may be compatible with one end of the DNA insert, while the other end of vector is compatible with the second end of the insert This situation arises when one end of the vector as well as that of the DNA insert is generated by one restriction enzyme, while their other ends are cut by a different enzyme. As earlier, the opened vector and DNA insert are mixed under annealing conditions to allow pairing between the compatible ends of the vector and the insert; T4 ligase is then used to seal the nicks to yield a recombinant DNA. *In such a situation, only the recombinant DNA is circularized (since the two ends of vector are*

unmatched as are the two ends of DNA insert) and the DNA insert is integrated in only one orientation or direction.

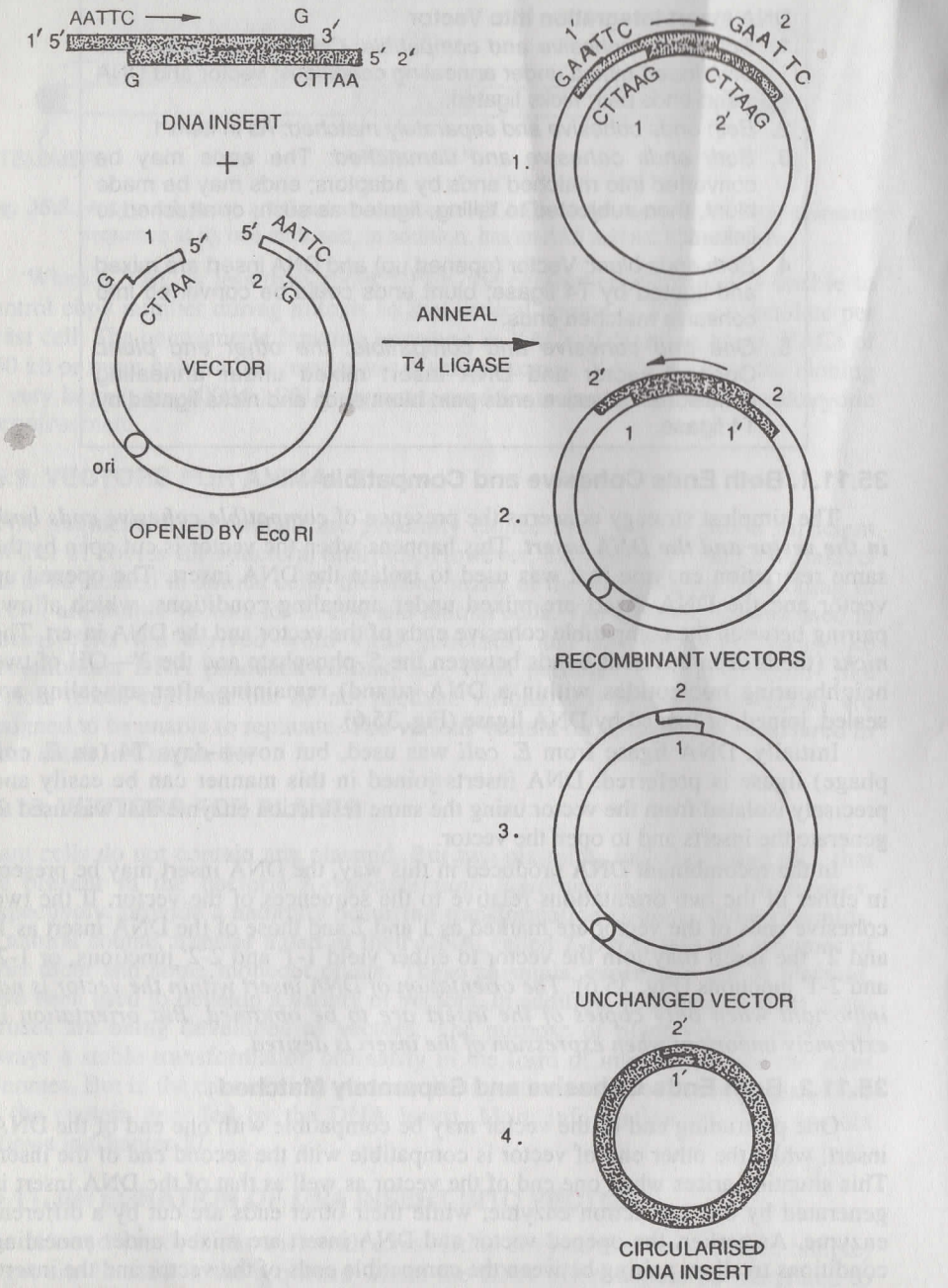


Fig. 35.6. The products of ligation when both the vector and the DNA insert are cut with the same restriction enzyme so that they have completely matched cohesive ends.

35.11.3. Both Ends Cohesive and Unmatched

Often the DNA insert is prepared using one restriction enzyme, while the vector is opened with another enzyme; this generates cohesive ends in the vector and the insert, which are unmatched. In this situation, the protruding ends may be converted into blunt ends, which may then be subjected to tailing to generate matched protruding ends. Alternatively linkers and/or adaptors may be employed to generate cohesive ends.

35.11.4. Both Ends Flush/Blunt

Flush or blunt-ended vector and DNA insert can be joined together by T4 DNA ligase. However, high concentrations of the enzyme, the vector and the DNA insert are required. The DNA insert gets ligated in either orientation, and it can be easily and precisely separated from the vector only if the two were cleaved by the same restriction enzyme (producing blunt ends). Alternatively, the blunt ends can be converted into cohesive ends as outlined above.

35.11.5. One End Cohesive and Compatible, the Other End Blunt

If the vector and the DNA insert have one compatible cohesive end and one flush end, their cohesive ends pair together when they are mixed under annealing conditions. The T_4 ligase seals the nick at the cohesive end and joins the blunt ends as well. It should be noted that under such a situation, only the recombinant DNA molecules are produced since vector and DNA insert molecules can not circularise due to their one cohesive and one blunt ends.

Transformation

- Introduction of DNA into a cell (here bacterial cell).
- *$CaCl_2$ treatment*: Cells suspended in 50 mM $CaCl_2$ and incubated for 12-24 hr; this treatment increases transformation frequency.
- The vectors that can be packaged in empty phage heads are used for infection of *E. coli* cells.

35.12. INTRODUCTION OF THE VECTOR INTO A SUITABLE HOST

The recombinant DNA is constructed *in vitro;* it is then generally introduced into *E. coli* to (1) select the recombinant DNA from the unchanged vector and (2) to obtain many copies of the recombinant DNA (DNA insert). Purified recombinant DNA may subsequently be introduced into another bacterium, yeast, higher plants or animals. The various approaches for introducing recombinant DNA into *E. coli* are briefly summarised here.

35.12.1. Increased Competence of *E. coli* by $CaCl_2$ Treatment

E. coli cells are generally poorly accessible to DNA molecules. But treatment with $CaCl_2$ makes them permeable to DNA; the process involved is poorly understood. Growing *E. coli* cells are isolated and suspended in 50 mM $CaCl_2$ at a concentration of 10^8 - 10^{10} cells/ml; the cells may be incubated for 12-24 hr to increase the frequency of transformation. The recombinant DNA is then added; efficient transformation takes only a few minutes and the cells are then plated onto a suitable medium for the selection of transformed clones.

The transformed cells are suitably diluted and spread thinly on a suitable medium so that each cell is well separated and produces a separate colony. Generally, the medium is so designed as to permit only the transformed cells to divide and produce colonies.

35.12.2. Infection by Vectors Packaged as Virions

Those vectors that have the λ phage *cos* sequences, *e.g.*, cosmids, phasmids and λ vectors, are generally packaged *in vitro* into specially produced empty λ phage heads and complete λ particles are constituted. These phage particles are used to infect *E. coli* cells; this is called ***transfection.*** These vectors can also be used to transform *E. coli* cells directly as naked DNA, using the $CaCl_2$ technique. Generally, infection by phage particles containing recombinant DNA is far more efficient than direct transformation.

The infected/transformed bacterial cells are spread on a lawn of susceptible cells, where clear areas or plaques develop in the lawn. Plaques containing the recombinant DNA (λ vector and phasmids containing DNA inserts) are identified and the phage particles collected from such plaques provide the purified recombinant DNA or vector DNA.

Selection of Bacterial Clones Having Recombinant DNA

- *Step 1*: Elimination of nontransformed clones.
- *Step 2*: Identification of clones having recombinant DNA.
- *Step 3:* Identification of clones having the desired DNA insert.

35.13. SELECTION OF RECOMBINANT CLONES

When recombinant DNA is constructed and used for transformation of *E. coli* cells, the following three types of bacterial cells are obtained: (1) a majority of the cells are nontransformed, (2) a proportion of the transformed cells contain the unaltered vector, while (3) the remaining cells have the recombinant DNA. The objective of cloning experiments is to identify and isolate those cells that contain the recombinant DNA containing the desired DNA insert. This is achieved in the following steps: (1) selection of clones containing the vector, (2) isolation of clones having recombinant DNA, and (3) identification of recombinant clones lowing the desired DNA insert.

35.13.1. Selection of Clones Containing Recombinant DNA

The first need is to eliminate the bulk of bacterial cells, which are non-transformed. This is generally achieved by inserting a selectable marker gene into the vector. A ***marker gene*** or ***reporter gene*** produces a phenotype, which permits either an easy selection or quick identification of the cells in which it is present. A ***selectable marker*** governs a feature, which enables only such cells that possess it to survive under the selective conditions. For example, genes conferring resistance to antimetabolites, *e.g.*, an antibiotic like kanamycin (Table 35.1), are good selectable markers. On the other hand, ***scorable markers*** produce distinct phenotypes, which allow an easy identification of the cells having them from those, which do not contain them.

35.13.1.1. Elimination of Nontransformed Cells

1. A good vector has at least two marker genes of which atleast one is selectable. The nontransformed bacterial cells are eliminated by plating them on a medium containing the concerned selection agent. All the colonies that develop on the selective medium are transformed either by the unaltered vector or the recombinant DNA.
2. In case of phage vectors, cosmids and phasmids, the vectors are packaged into virions and used for infection of bacterial cells. The infected cells are plated on a susceptible bacterial lawn where they form plaques. A plaque so produced may contain either recombinant DNA or unchanged vector.

35.13.1.2. Identification of Clones Having Recombinant Vectors

The second step consists of identification and isolation of those clones that are transformed by the recombinant DNA from among those that contain the unaltered vector. This may be achieved in one of the several ways listed below.

1. In case the vector has two selectable markers, *e.g.*, *p*BR322, the strategy described in section 35.4.1 may be followed.
2. When a vector has a selectable and a scorable marker, the strategy outlined in Section 35.5.2 may be employed.
3. When the DNA insert itself codes for a gene product, which is defective in the auxotrophic host cells, a direct selection for the recombinant DNA is possible. The host cells are grown on a medium lacking the compound needed by the auxotrophic host; only those cells which contain the recombinant DNA can grow and form colonies.
4. The selection strategies for λ vectors are outlined in section 35.5.2.

Selection of Clones Having the Desired Insert

- *Colony hybridization*; uses radioactive probes specific for the insert.
- *Hybrid-arrested translation*; *in vitro* translation system.
- *Hybrid selection*; uses *in vitro* translation system.
- *Complementation.*
- *Antibodies* specific to DNA insert encoded protein; ELISA is the most sensitive technique.
- Fluorescence activated cell sorting (FACS).

35.13.2. Selection of the Clone Containing A Specific DNA Insert

Once we obtain a population of recombinant colonies, the next step is to identify a clone that has the DNA insert of interest. The various strategies used for the purpose are briefly outlined below.

35.13.2.1. Colony Hybridization

The most efficient and rapid strategy for identification of a clone having the desired insert uses the technique of colony hybridization. The bacterial colonies (containing recombinant DNA) are replica-plated or phage plaques are directly lifted on to nitrocellulose filters, the cells are lysed and their DNA is denatured. This filter is incubated with the specific radioactive (^{32}P-labelled) probe under annealing conditions. After some time, the probe is washed out leaving only those probe molecules that have hybridized with the denatured DNA from bacterial cells or

phage particles. The colonies/plaques with whose DNA the probe has hybridized are identified by autoradiography; these contain the desired DNA insert. These colonies/plaques are isolated from the master plate used for replica-plating.

It is essential that a probe specific for the desired DNA insert is available. A ***probe*** is a polynucleotide (DNA or RNA, usually of 25-30 bases) molecule of a specific base sequence, which is used to detect DNA molecules containing the same base sequence. Generally, the probes are labelled with ^{32}P to enable auto-radiography for an easy identification of the DNA samples that base-pair with the probe. It is desirable that the probes are single-stranded to avoid pairing between the two strands of the probe itself.

35.13.2.2. Other Approaches

When specific probes are not available, many indirect approaches may be used for the identification of clones having the desired DNA insert. These procedures are not generally convenient for screening of a large number of clones. Two of such procedures, called (1) hybrid-arrested translation (HART) and (2) hybrid selection, use *in vitro* translation systems and the identification of resulting polypeptide(s). *It is, therefore, necessary that the protein product of the DNA insert being searched should be known, at least in terms of its electrophoretic mobility.*

35.13.2.3. Complementation

The cloned DNA insert may express itself in the bacterial cells, *e.g.*, prokaryotic genes, some yeast genes and eukaryotic cDNAs. Eukaryotic sequences isolated from genomic DNA have to be expressed in appropriate eukaryotic hosts, *e.g.*, yeast cells, animal cells in culture, etc. If the protein produced by the desired DNA insert is deficient in the host cells, this DNA insert will correct the deficiency of the transformed cells, *i.e.*, will ***complement*** the deficiency of host cells. This strategy is limited in application by the availability of appropriate host cells.

35.13.2.4. Antibodies Specific to the Protein Product

A protein can be detected by using antibodies specific to it. A practical approach is to divide the large number of recombinant clones into a convenient number of groups and to assay for the presence of the protein. The positive group is again divided into subgroups and assayed. In this manner, the positive groups are subdivided again and again till a single positive clone is identified. The identification of proteins using antibodies may be done by western blotting, precipitation and electrophoresis or ELISA (enzyme- linked immunosorbent assay).

35.13.2.5. FACS

In case of animal cells, an automated system, called ***fluorescence activated cell sorter (FACS),*** can be used for a very rapid (upto 1,000 cells/sec) sorting of transformed cells. This is applicable to all the genes whose products become arranged on the cell surface and are available for binding of specific antibodies. Therefore, these proteins must not be produced by the nontransformed host cells. The antibody molecules are attached to a fluorescent molecule and the transformed cells are treated with this antibody specific for the desired protein. The cells containing on their surface the protein in question will bind with the fluorescent antibodies. Cells are then passed one by one in a stream between a laser beam and a

fluorescence detector. The cells that fluoresce are deflected into a microculture tray, while the nonfluorescing cells are drawn away by an aspirator.

35.14. MULTIPLICATION EXPRESSION AND INTEGRATION OF THE DNA INSERT IN HOST GENOME

Once the *E. coli* clone containing the desired DNA insert is identified, it is multiplied to obtain sufficient number of copies of the DNA insert to be used in one or more of the following ways. (1) It can be used for a structural analysis of the insert, *e.g.*, DNA sequencing, chromosome walking, etc. (2) It may be introduced into a bacterium like *B. subtilis* for production of the protein encoded by the insert since this host secretes proteins into the medium, which allows easy purification. (3) It can be introduced into a eukaryotic host, *e.g.*, yeast, animal cells, plants, etc., either to investigate the function of the insert or (4) to integrate it into the host genome to achieve one of many diverse objectives.

QUESTIONS

1. Define vector. Briefly describe the various kinds of vectors for *E. coli.*
2. What are the features of a good vector ? Briefly describe the features of one plasmid and one phage vector of *E. coli.*
3. Describe in some detail the various strategies for the integration of DNA inserts into the vector, and for selection of cells carrying the recombinant DNA.
4. Explain the various strategies used for selection of clones having recombinant DNAs and those for isolation of colonies containing the desired DNA insert.
5. Briefly describe the different strategies used for genetic transformation of *E. coli,* and for the selection of recombinant clones.
6. Write short notes on the following:
   (*i*) cosmid, (*ii*) colony hybridization,
   (*iii*) phage vectors, (*iv*) shuttle vectors,
   (*v*) expression vectors, (*vi*) YAC,
   (*vii*) transformation, (*viii*) integration of DNA insert into vectors,
   (*ix*) selection of recombinant clones.

36

Transgenic Animals

36.1. INTRODUCTION

A ***transgenic animal*** contains in its genome, a gene or genes introduced by one or the other technique of transfection. The gene introduced by transfection is called ***transgene.*** In animals, ***transfection*** specifies the introduction of a DNA segment, either naked or integrated into a vector, into an animal cell. It may be pointed out that this phenomenon is known as ***transformation*** in all other organisms. However, in case of animals, ***transformation*** has long been used to describe the change of normal, *i.e.,* nontumourous, cells in culture to tumour like cells. Therefore, it became necessary to use the term transfection to avoid confusion.

Transfection may be transient or permanent (stable). In ***transient transfection,*** the introduced gene(s) is(are) gradually lost from the daughter cells of transfected cells. But in case of ***stable transfection,*** the introduced gene(s) is(are) retained and expressed in all the cells derived from the transfected cells. Since most of the animal vectors are unstable, *i.e.,* are gradually lost, in the extra chromosomal state, stable transfections are ordinarily due to the integration of introduced gene(s) into the cell genome.

36.2. OBJECTIVES OF GENE TRANSFER

The various objectives for which transgenic cells/animals are produced may be summarised into the following five categories.

1. Genes have been transferred and expressed into cultured cell lines to obtain the proteins encoded by them.
2. Genetic modification of animals may be aimed at improving their milk, meat, wool, etc. production.
3. Genes have been transferred into animals with a view to obtain a large scale production of the proteins encoded by these genes in the milk, urine or blood of such animals. Such animals are often called ***bioreactors*** and the approach is referred to as ***molecular farming*** or even as ***gene farming.***
4. A special case of gene transfer aims at alleviating or even eliminating the symptoms and consequent miseries of genetic diseases; this is called ***gene theray.***
5. Finally, specific transgenic animal strains or lines are created to fulfil specialized experimental and/or biomedical needs, *e.g.,* "knockout" mice strains in which specific genes have been disrupted.

Transfection
- Introduced of DNA into animal cells.
- Two: types
  1. *Transient* : Introduced gene gradually lost from the cell population.
  2. *Stable* : Introduced genes retained in all cells derived from the transfected cell; ordinarily due to transgene integration into the cell genome.
- *Transgene* : A gene introduced into an organism by genetic engineering.

Objectives of Gene Transfer
- Protein production from cell cultures.
- Enhanced production or quality of transmnic animals.
- Protein production from transgenic animals.
- Gene therapy
- Production of specific experimental animals, *e.g.,* 'knockout' mice.

Vectors
- Derived from viruses, retroviruses and transposable elements.
- *Types* :
  A. Ability to infect host cells:
  1. *Virus-like* : Produce virions that infect host cells.
  2. *Plasmid-like*: Behave like bacterial plasmids.
  B. Ability to replicate in host cells:
  1. *Replicating vectors*: replicate in host cells; most vectors.
  2. *Non-replicating vectors*: Unable to replicate in host cells; used to deliver DNA for integration into the host genome.

36.3. VECTORS

The various animal vectors are based on one or the other virus, *e.g.,* SV40, bovine papillamovirus, retroviruses, etc., or on a transposable element, *e.g., Drosophila P* element (Table 36.1). Some of these vectors behave like viruses in that they produce virions, which are used to infect the host cells. Some other vectors are like bacterial plasmids; these vectors have to be introduced into the cells using a suitable transfection technique. Most of the vectors are ***replicating vectors*** in that they have the ability to replicate in suitable animal host cells. But some of them are ***non-replicating vectors***; these are used to carry DNA sequences into the host cells for integration into the animal genomes.

Any DNA fragment, or a mixture of fragments can be used to transfect animal cells; these DNA segments become integrated into the cellular genome and express themselves. Further, vectors that remain free and replicate in extra-chromosomal state are usually quickly lost in a few days even when selection conditions favoring their retention are imposed; this is called ***transient transfection. Permanent*** or ***stable transfection*** most often is due to the integration of transfecting DNA into the genome of animal cells; this is also known as ***insertional transfection.*** *DNA integration is usually by nonhomologous recombination, hence in random locations in the genome.*

The first animal vector was derived from the primate papova virus, simian virus 40 (SV40). But subsequently, vectors have been developed from many other viruses, *e.g.*, papilloma virus, adenoviruses, the Epstein-Barr herpes virus, vaccinia viruses (all for mammals) and baculoviruses (for insects). SV40 genome has been used to develop mainly 3 types of vectors: (1) transducing vectors, (2) plasmid vectors and (3) transforming vectors.

SV40 Vectors
- Based on simians virus 40 (SV40) genome.
- *Types:*
  1. Transducing vectors.
  2. Late replacement vectors.
  3. Transforming vectors.

36.3.1. SV40 Transducing Vectors

These vectors produce viral particles after infecting monkey cells. They must have the following three features: (1) the SV40 origin including the surrounding region containing the transcriptional regulatory signals (*i.e.*, regions at which splicing and polyadenylation occur), (2) size (including, that of the DNA insert) between 3900 bp and 5300 bp for packaging into virions, and (3) genes encoding large-T, VP1, VP2, and VP3. This leaves very little room for DNA inserts. But the genes encoding the necessary proteins, *viz.*, larger-T, VP1, VP2 and VP3, can be present in another virus or within the host genome. This flexibility is used to advantage for solving the size and selection problems.

SV40 Transducing Vector
- Produce virions in monkey cells; virus - like vector.
- Features:
  1. SV40 origin + regulatory sequences.
  2. Size between 34 and 53 kb.
  3. Gene for large-T, VP1-VP3.
- Restriction on DNA insert size.

36.3.2. Late Replacement Vectors

The region encoding VP1, VP2 and VP3 may be replaced in the vector by DNA insert; such a vector is called ***late region replacement vector,*** *e.g.*, SVGT-5 (an improved SV40 vector). A vector of this type is used for infection of host cells in conjunction with another virus, called ***helper virus,*** which has the VP1, VP2, and VP3 genes intact but has a defective large-T gene. In this case, only those host cells that are infected by both the vector and the helper virus will lyse and produce virions since cells infected by either the vector or the helper virus alone will not support packaging or replication (respectively) of the virus. This feature is very useful since all the plaques formed on monkey cell monolayers will contain the vector.

36.3.3. Early Replacement Vectors

Alternatively, the essential genes missing from the vector may be present within the genome of host cells. For example, COS (CV-1, Origin of SV40; CV-1 is a monkey cell line) cell line of African green monkey kidney cell cultures contains in its genome the gene for large-T of SV40. Therefore, a vector having the origin of

Late Replacement SV40 Vectors
- Genes encoding VP1, VP2 and VP3 are replaced by DNA insert.
- Used to infect host cells together with a 'helper' virus.
- *Helper Virus* has intact VP1-VP3 genes but defective large-T gene.
- All the plaques formed on monkey cell monolayer have the vector

replication and genes for VP1, VP2 and VP3 will replicate and produce virions in COS cell line cells. In such a case, no helper virus is required. Since in such a vector the early genes (large-T gene) are replaced by the DNA insert, it is called ***early region replacement vector.***

Early Replacement Vector
- Vector lacks large-T gene; a virus-like vector.
- Large-T gene present in host cell genome, *e.g.*, in COS cell line.
- Helper virus not required.

36.3.4. SV40 Plasmid Vectors

These vectors replicate in monkey cells but do not get packaged into virions. They contain the origin of replication and the large-T encoding gene (large-T gene is not necessary for multiplication in COS cells). Obviously, there is no size limit on

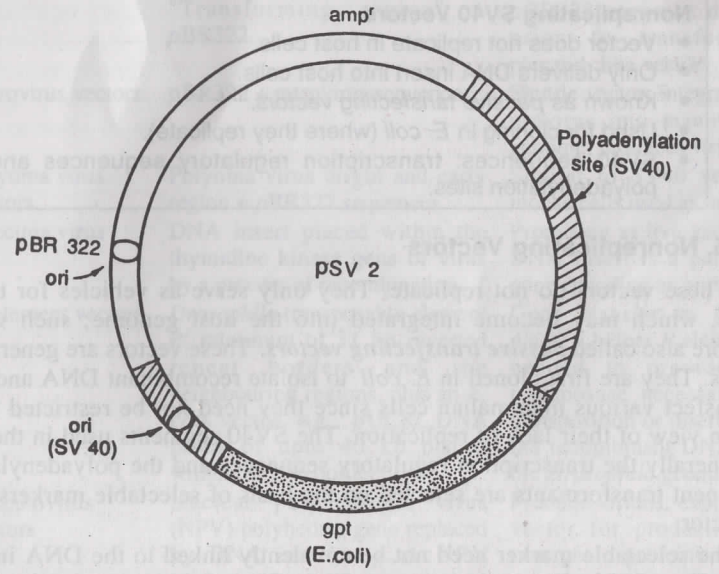


Fig. 36.1. A SV40-based plasmid vector. It is a shuttle vector having origin and *amp^r* gene from *p*BR322 (for function in *E. coli*), and the following SV40 segments: the origin and the neighbouring trnscriptional control sequences including early promoter, and the polyadenylation site. This vector functions as a plasmid vector in host cells having the large-T gene, *e.g.*, COS cell line, but as a nonreplicative vector in other host cells.

such vectors, and some of them are *E. coli* and monkey shuttle vectors, *e.g.*, *p*SV2, *p*SV3, etc. (Fig. 36.1). These vectors produce high copy number per cell. The shuttle vectors are used to propagate the recombinant DNA in *E. coli*, which is then introduced into monkey cells to study the expression of DNA inserts.

SV40 Plasmid Vector
- Vector behaves like plasmid.
- It contains:
  1. SV40 origin,
  2. large-T gene, and
  3. other necessary functions.
- No size limit on vectors.
- Some are *E. coli* shuttle vectors.
- Generally, used for transient transfection only.
- Integrates in host genome with frequency of 10^{-5}-10^{-3}.

Plasmid vectors are unstable in monkey cells, and are generally used for transient transfection only. These vectors can be stably maintained if the large-T function is provided by COS cells and the vector has a selectable marker, say, *E. coli neo* gene, and the host cells are maintained under the selection environment, in this case on G-418. Plasmid vectors also integrate into the host cell genome with a frequency of 10^{-5} to 10^{-3}; this yields stable transfection.

Nonreplicating SV40 Vectors
- Vector does not replicate in host cells.
- Only delivers DNA insert into host cells.
- Known as *passive tansfecting vectors.*
- Used for cloning in *E. coli* (where they replicate).
- SV40 sequences: transcription regulatory sequences and polyadenylation sites.

36.3.5. Nonreplicating Vectors

These vectors do not replicate. They only serve as vehicles for transfecting DNAs, which may become integrated into the host genome; such vectors are therefore also called ***passive transfecting vectors.*** These vectors are generally shuttle vectors. They are first cloned in *E. coli* to isolate recombinant DNA and then used to transfect various mammalian cells since they need not be restricted to monkey cells in view of their lack of replication. The SV40 segments used in these vectors are generally the transcription regulatory sequences and the polyadenylation sites. Permanent transformants are selected on the basis of selectable markers present in the vector.

The selectable marker need not be covalently linked to the DNA insert or the transfecting DNA. Even when two separate DNA fragments containing separate genes are mixed and used for transfection, more than 50% of the permanently transfected cells contain both the genes, usually integrated side by side. The two DNA fragments tend to become joined after entering the animal cells, which is the reason for their ***co-transfection***, *i.e.*, integration of the two genes together in the genome.

TABLE 36.1
The different type of vectors used for gene transfers in animals

Vector	Derived from	Features
1. SV40 vector		
(*a*) Early region replacement vectors	Replacement of large-T gene of SV40	1. Produce virions, which infect host cells 2. Transient gene expression 3. Mammalian cells are hosts of SV40
(*b*) Late region replacement vectors	Replacement of VP1, VP2, and VP3 genes of SV40, *e.g*, SVGT-5	
(*c*) Plasmid vectors	Origin of replication and large- T gene of SV40	—
(*d*) Shuttle plasmid vectors	Plasmid vector [item 1(*c*)] plus *p*BR322 origin and *amp*^{r} gene, *e.g.*, *p*SV2, *p*SV3, etc.	—
(*e*) Passive transfecting vectors	Rous sarcoma virus promoter in place of SV40 early promoter, *e.g.*, *p*RSV.	Strong expression of the marker gene
	SV40 transcription regulatory and polyadenylation sequences, plus *p*BR322 origin and *amp*^{r} gene	Shuttle vectors; used for gene integration in mammalian cells
2. BPV vectors	Bovine papilloma virus "Transforming region" + *p*BR322.	Shuttle vector; often *p*BR322 sequence deleted prior to transfection; plasmid- like vector
3. Retrovirus vectors	*p*BR322 + retrovirus sequences	Shuttle vector; integrates as provirus into mammalian genome; produces virions
4. Polyoma virus vectors	Polyoma virus origin and early region + *p*BR322 sequences	Similar to SV40 vectors; mouse cells used as host.
5. Vaccinia virus	DNA insert placed within the thymidine kinase gene of virus by a process of recombination.	Promising as live vaccines; DNA insert is a pathogen gene encoding an antigen.
6. *P* element vectors	*Drosophila* transposable element P; minimum of 31 bp inverted repeat borders and the neighbouring regions, plus an *E. coli* vector, *e.g.*, *p*UC8; DNA insert of upto 40 kb placed within the two borders.	Gene transfer in *Drosophila*; a helper *P* element is needed to provide the transposase necessary for transposition or insertion of the recombinant DNA into the *Drosophila* genome
7. Bacculovirus vectors	Nuclear polyhedroma virus (NPV) polyhedrin gene replaced by DNA insert; *e.g.*, AcNPV (*Autographa californica* nuclear polyhedroma virus) and BmNPV (*Bombyx mori* nuclear polyhedroma virus) vectors	Produce virions; expression vector for production of transgenic proteins in silk worm larvae (BmNPV vectors) and in *Spodoptera frugiperda* larvae or cultured cells (AcNPV vectors)

Co-transfection

- When two separate DNA fragments are introduced in a cell, more than 50% of the permanently transfected cells contain both the DNA segments.
- The two DNA segments are integrated at the same site in tandem.
- The two DNA segments become linked end - to - end before integration.

36.4. GENE CONSTRUCTS

The purpose of gene transfer may range from study of promoter function, expression of reporter genes, through transcription and translation, products of any known or unknown DNA sequence of interest, regulation of gene expression, actions of oncogenes, to the identification of transcription and translation factors, etc. Many of these molecular studies are conveniently performed using *Xenopus laevis* oocytes. The approach to carry out genetic analyses using an organism or cell different from that whose genes are being studied has been termed as ***surrogate genetics.***

A transgene must be integrated into the host genome for obtaining transgenic cells/animals, while for molecular biology studies the transgenes are ordinarily present in an extrachromosomal state. But in either case, the transgene must be present in proper orientation in relation to and in association with the various sequences required for its efficient transcription and translation in the host cells. The various sequences included in the gene construct may be listed as follows: (1) an efficient promoter/enhancer, (2) the translation initiation codon (AUG in mRNA, and ATG in DNA), (3) the chain termination codon(s), (4) transcription termination sequence, including (5) polyadenylation cleavage/addition site (Fig. 36.2). A suitable expression vector ordinarily contains a promoter, transcription termination

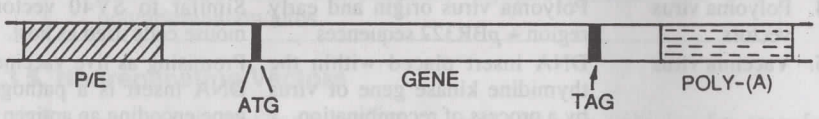


Fig. 36.2. A simplified schematic representation of a gene construct for expression in animal/plant cells. P/E, promoter/enhancer sequence; ATG (= AUG), the initiation codon; Gene, the gene (may contain introns) that is to be expressed; Te, chain termination signal (during translation); Poly(A), transcription termination including poly(A) additionl/deletion site. Suitable signal sequences must be included for protein transaction, if required.

sequence and, usually, a polyadenylations site, *e.g.*, *p*SV vectors (Fig. 36.1). Further, some marker genes have to be used for identification and selection of the transfected cells; such genes are called ***reporter genes*** (Fig. 36.3). In addition, the gene must contain appropriate signalling sequences for the transport, etc. of the protein encoded by it.

36.4.1. Promoters

A ***promoter sequence*** is the site to which RNA polymerase first binds during the initiation of transcription. A promoter sequence is absolutely essential for the

Promoters

- Site for RNA polymerase binding; essential for transcription.
- Determines:
  1. Cell/tissue etc. where transcription occurs.
  2. The level of transcription.
  3. Whether expression is constitutive or tissue, etc. specific.
- *Enhancers*: A sequence that enhances activity of a promoter.
- Promoter may be used in conjunction with an enhancer.

transcription of a DNA segment. The promoter must be located at that end of the gene, which has the initiation codon AUG, *e.g.*, upstream of the coding region of the gene. In contrast, an ***enhancer sequence*** is itself not the site of RNA polymerase binding, but it enhances the activity of promoter located in its neighbourhood often upto several kilobases away.

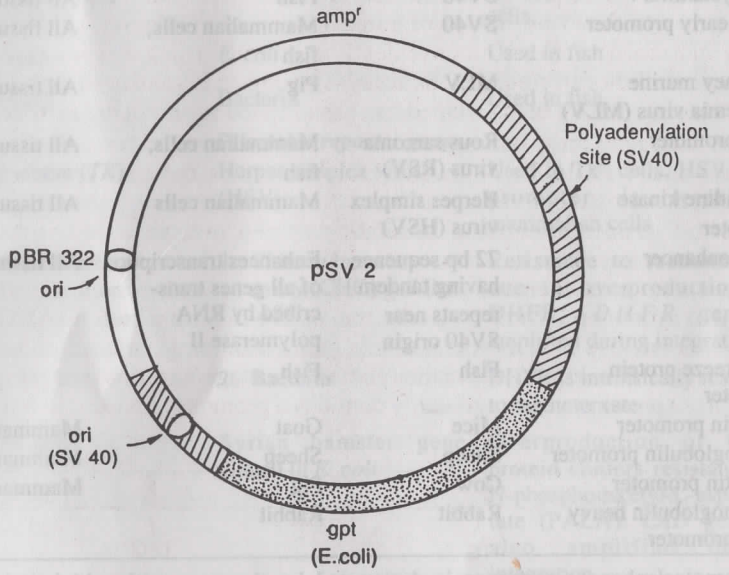


Fig. 36.3. A schematic representation of a plasmid gene construct having a selectable marker (*amp*^{r}) gene. This construct is a vector prototype for fish. *amp*^{r}, ampicillin resistance; P/E, promoter/enhancer; Poly(A), polyadenylation site; and *ori*, origin of DNA replication.

The promoter sequence determines not only the level of transcription of a gene but also the tissue or cell type, the developmental stage, etc. at which gene expression will take place. For example, SV40 early promoter is a constitutive promoter, *i.e.*, is expressed in a wide variety of mammalian cells. Similarly, RSV (Rous sarcoma virus) promoter is also constitutive, but it is much more powerful than the SV40 promoter. In contrast, β-lactoglobulin promoter/enhancer is a mammary gland tissue- specific promoter so that genes fused with this promoter are expressed only in the mammary glands. A great deal of thought and care goes into the selection of a suitable promoter to be used in a gene construct (Table 36.2). A suitable enhancer sequence may also be used along with the promoter either to

enhance the promoter activity or to achieve a degree of specificity in terms of the tissue in or the conditions under which the gene is to be expressed.

TABLE 36.2
A list of promoter/enhancer sequences used for driving transgenes in different trasgenic animal species

*Promoter/enhancer sequence**	*Source*	*Used in animal species*	*Gene expression in the tissue*
Metallathionein promoter	1. Fish	Fish	
	2. Mouse	Fish, mouse, pig	
	3. Sheep	Sheep	
	4. Human	Rabbit, pig	
β- actin	1. Fish	Fish	
	2. Chicken	Fish	
cd- crystallin SV	SV40	Fish	All tissues
SV40 early promoter	SV40	Mammalian cells, fish	All tissues
Moloney murine leukaemia virus (MLV)	MLV	Pig	All tissues
RSV promoter	Rous sarcoma virus (RSV)	Mammalian cells, fish	All tissues
Thymidine kinase promoter	Herpes simplex virus (HSV)	Mammalian cells	All tissues
SV40 enhancer	72 bp sequence having tandem repeats near SV40 origin	Enhances transcription of all genes transcribed by RNA polymerase II	All tissues
Anti-freeze protein promoter	Fish	Fish	
β-casein promoter	Mice	Goat	Mammary tissue
β-lactoglobulin promoter	Sheep	Sheep	Mammary tissue
Prolactin promoter	Cow	Pig	Mammary tissue
Immunoglobulin heavy chain promoter	Rabbit	Rabbit	

*A promoter/enhancer sequence is designated by the gene with which it is naturally associated.

Reporter Gene
- *Scorable Markers*:
  1. *CAT* (chloramphemicol acetyltransferase)
  2. B-galactosidase.
  3. *lux* (heciferase).
- *Selectable Markers*:
  1. Thymidine kinase (*TK*).
  2. Dihydrofolate reductase (*DHFR*).
  3. CAD Protein gene.
  4. Xanthine - guanine phosphoribosyl transferase (*XGPRT*).
  5. Neomycin phosphotransferase (*neo*).

36.4.2. Reporter or Marker Genes

Scorable markers are used for studying promoter/enhancer activity, while selectable markers are utilized to select for transfected cells. Examples of scorable makers are chloramphenicol acetyltransferase (CAT) gene from *E. coli* transposon Tn9, β-galactosidase gene from *E. coli* and bacterial luciferase gene. The selectable reporter genes commonly used are thymidine kinase (*TK*), dihydrofolate reductase (methotrexate resistance), CDA protein (PALA resistance), XGPRT (mycophenolic acid resistance), and neomycin phosphotransferase (G418 resistance) (Table 36.3).

TABLE 36.3
A list of reporter or marker genes used in gene transfers in animal cells

Reporter gene	*Source*	*Remarks*
	Scorable reporter genes	
Chloramphenical acetyl transferase (*CAT*)	*E. coli* transposon Tn9	Used in fish, mammalian cells,
β-galactosidase	*E. coli*	Used in fish
Luciferase (*lux*)	Bacteria	Used in fish
	Selectable reporter genes	
Thymidine kinase (*TK*)	Herpes simplex virus (HSV)	Used in Tk^- cells; HSV gene promoter is weak in mammalian cells
Dihydrofolate reductase (*DHFR*)	1. Chinese hamster ovary (CHO) cell line 'A29'	Resistance to methotrexate due to overproduction of DHFR; *DHFR* gene is amplified during integration
	2. Bacteria	DHFR is intrinsically resistant to methoterxate
CAD protein gene	Syrian hamster gene cloned in *E. coli*	Overproduction of CAD protein confers resistance to N-phosphonacetyl-L-aspartate (PALA). *CAD* gene is also amplified during integration
Xanthine- guanine phosphoribosyltransferase (*XGPRT*)	*E. coli*	Resistance to mycophenolic acid
Neomycin phosphotransferase (*neo*)	Bacterial transposons Tn5 and Tn601	Resistance to antibiotic G418

Transfaction Methods
1. Calcium phosphate precipitation.
2. Microinjection.
3. Retrovirus infection.
4. Lipofection.
5. Particle gun.
6. Electroporation.

36.5. TRANSFECTION METHODS

Several approaches have been used for the introduction of DNA into animal cells/embryos, which are as follows: (*i*) calcium phosphate precipitation (Fig. 36.4), (*ii*) direct microinjection (Fig. 36.5), (*iii*) retrovirus infection, (*iv*) lipofection, (*v*) particle gun delivery and (*vi*) electroporation. The important strategies are briefly described below.

Calcium Phosphate Precipitation
- DNA dissolved in phosphate buffer.
- $CaCl_2$ is then added.
- Calcium phosphate co-precipitates with DNA.
- The precipitate is added to the cultured cells.
- Cells take up DNA by phagocytosis.
- Upto 20% cells transfected.
- Applicable to virtually all mammalian cell cultures.

36.5.1. Calcium Phosphate Precipitation

In this approach, the DNA preparation to be used for transfection is first dissolved in a phosphate buffer. Calcium chloride solution is then added to the DNA solution; this leads to the formation of insoluble calcium phosphate, which co-precipitates with the DNA. The calcium phosphate-DNA precipitate is added to the cells to be transfected. The precipitate particles are taken in by the cells by phagocytosis (Fig. 36.4). Transfection of upto 20% of the cells can be achieved. In a small proportion of the transfected cells, the DNA becomes integrated into the cell genome producing stable or permanent transfection. This general approach can be applied to virtually all mammalian cells, and a very large number of cells can be treated with little efforts. But many cell lines do not like the calcium phosphate precipitate adhering to their surface.

This procedure is highly suited for transient transfection used for various molecular biology studies, particularly using COS cell lines. However, again for some unknown reason, it is not efficient in producing stable transfection.

Lipofection
- DNA delivery through liposomes.
- *Liposome*: A small vesicle prepared from a suitable lipid.
- *Liposomes from nonionic lipids*: special procedures required to place DNA within liposomes.
- *Liposomes from cationic lipids*:
  1. DNA binds on the outside of liposomes by electrostatic attraction.
  2. Fuse with plasma membrane.
  3. Deliver DNA inside cell.
- Method of choice for mammalian cells *in vitro*.
- Used for DNA delivery in live animals.

36.5.2. Lipofection

The delivery of DNA into cells using liposomes is called ***lipofection***. ***Liposomes*** are small vesicles prepared from a suitable lipid. Initially, nonionic lipids were used for preparing liposomes so that DNA had to be introduced within the vesicles following specific encapsidation procedures. The use of cationic lipids for the construction of liposomes is a distinct advantage as DNA spontaneously and efficiently complexes with these liposomes making encapsidation procedures unnecessary.

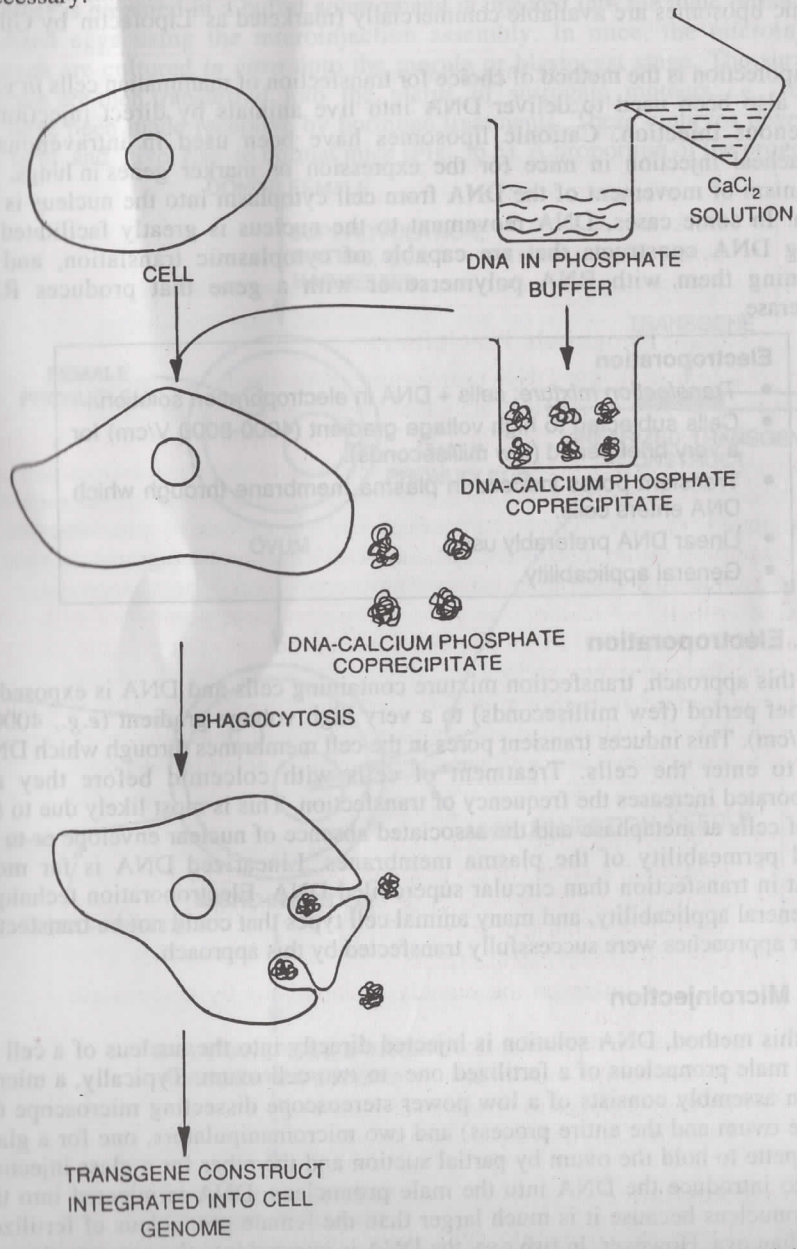


Fig. 36.4. A schematic representation of transfection by calcium phosphate precipitation.

Cationic liposomes have a single lipid bilayer membrane (unilamellar), and they bind to the cells efficiently. Probably they fuse with the plasma membrane and thereby deliver the DNA into the cells, which brings about transfection. DNA binds on the outside by electrostatic attraction. These liposomes cause perturbations in plasma membrane due to which they fuse and the DNA enters into the cytoplasm. Cationic liposomes are available commercially (marketed as 'Lipofectin' by Gibco-BRL).

Lipofection is the method of choice for transfection of mammalian cells *in vitro*. It has also been used to deliver DNA into live animals by direct injection or intravenous injection. Cationic liposomes have been used in intravenous or intratracheal injection in mice for the expression or marker genes in lungs. The mechanism of movement of the DNA from cell cytoplasm into the nucleus is not known. In some cases, DNA movement to the nucleus is greatly facilitated by making DNA constructs that are capable of cytoplasmic translation, and by combining them with RNA polymerse or with a gene that produces RNA polymerase.

Electroporation

- *Transfection mixture*: cells + DNA in electroporation solution.
- Cells subjected to high voltage gradient (4000-8000 V/cm) for a very brief period (few milliseconds).
- Transient pores induced in plasma membrane through which DNA enters cells.
- Linear DNA preferably used.
- General applicability.

36.5.3. Electroporation

In this approach, transfection mixture containing cells and DNA is exposed to very brief period (few milliseconds) to a very high voltage gradient (*e.g.*, 4000 - 8000 V/cm). This induces transient pores in the cell membranes through which DNA seems to enter the cells. Treatment of cells with colcemid before they are electroporated increases the frequency of transfection. This is most likely due to the arrest of cells at metaphase and the associated absence of nuclear envelope or to an unusual permeability of the plasma membranes. Linearized DNA is far more efficient in transfection than circular supercoiled DNA. Electroporation technique has a general applicability, and many animal cell types that could not be transfected by other approaches were successfully transfected by this approach.

36.5.4. Microinjection

In this method, DNA solution is injected directly into the nucleus of a cell or into the male pronucleus of a fertilized one- to two-cell ovum. Typically, a micro-injection assembly consists of a low power stereoscope dissecting microscope (to view the ovum and the entire process) and two micromanipulators, one for a glass micropipette to hold the ovum by partial suction and the other for a glass injection needle to introduce the DNA into the male pronucleus. DNA is injected into the male pronucleus because it is much larger than the female pronucleus of fertilized mammalian ova. However, in fish ova, the DNA is injected into the egg cytoplasm.

The general procedure for microinjection is as follows (Fig. 36.5). Donor females are induced to superovulate using appropriate hormone treatments. The superovulated females are then mated with fertile males, and large numbers of fertilized one- to two-cell ova/embryos are collected surgically. The transgene construct is prepared in a buffer solution and is injected into the male pronuclei of fertilized eggs using the microinjection assembly. In mice, the microinjected embryos are cultured *in vitro* upto the morula or blastocyst stage. The surviving embryos are then transferred into the uterus of surrogate mothers, *i.e.*, females that have been made receptive by hormone treatments. These embryos develop to full term and give rise to normal mice (Fig. 36.5). A proportion of the progeny so

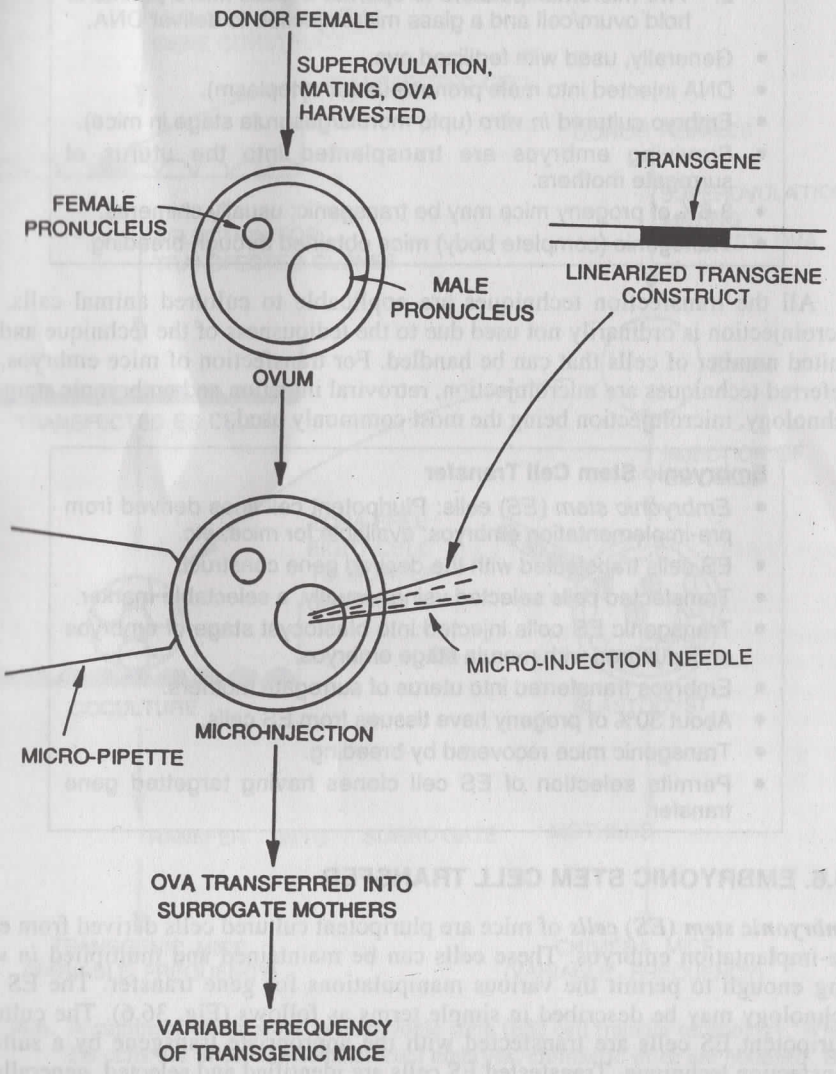


Fig. 36.5. A schematic representation of the microinjection technique of transfection for producing transgenic mice.

produced will be transgenic in that all their cells will contain the transgene stably integrated into their genomes. In mice, an average of about 3-6% of the progeny derived from microinjected embryos are transgenic; the frequency is much lower in other animals, *e.g.*, < 1% in sheep and pigs.

Microinjection

- DNA injected directly into the nucleus of a cell.
- *Microinjection assembly:*
  1. A stereoscopic microscope.
  2. Two micromanipulators to operate a glass micro-pipette to hold ovum/cell and a glass microinjection to deliver DNA.
- Generally, used with fertilized ova.
- DNA injected into male pronucleus (or cytoplasm).
- Embryo cultured *in vitro* (upto morula/gasgrula stage in mice).
- Surviving embryos are transplanted into the uterus of surrogate mothers.
- 3-6% of progeny mice may be transgenic; usually chimeras.
- Transgenic (complete body) mice obtained through breeding.

All the transfection techniques are applicable to cultured animal calls, but microinjection is ordinarily not used due to the tediousness of the technique and the limited number of cells that can be handled. For transfection of mice embryos, the preferred techniques are microinjection, retroviral infection and embryonic stem cell technology, microinjection being the most commonly used.

Embryonic Stem Cell Transfer

- *Embryonic stem* (*ES*) cells: Pluripotent cell lines derived from pre-implementation embryos; available for mice, etc.
- ES cells transfected with the desired gene construct.
- Transfected cells selected using, usually, a selectable marker.
- Transgenic ES cells injected into blastocyst stage of embryos or cocultured with morula stage embryos.
- Embryos transferred into uterus of surrogate mothers.
- About 30% of progeny have tissues from ES cells.
- Transgenic mice recovered by breeding.
- Permits selection of ES cell clones having targetted gene transfer.

36.6. EMBRYONIC STEM CELL TRANSFER

Embryonic stem* (*ES*) *cells of mice are pluripotent cultured cells derived from early pre-implantation embryos. These cells can be maintained and multiplied *in vitro* long enough to permit the various manipulations for gene transfer. The ES cell technology may be described in simple terms as follows (Fig. 36.6). The cultured piuripotent ES cells are transfected with the appropriate transgene by a suitable transfection technique. Transfected ES cells are identified and selected, generally by employing a selectable marker gene, and are cloned.

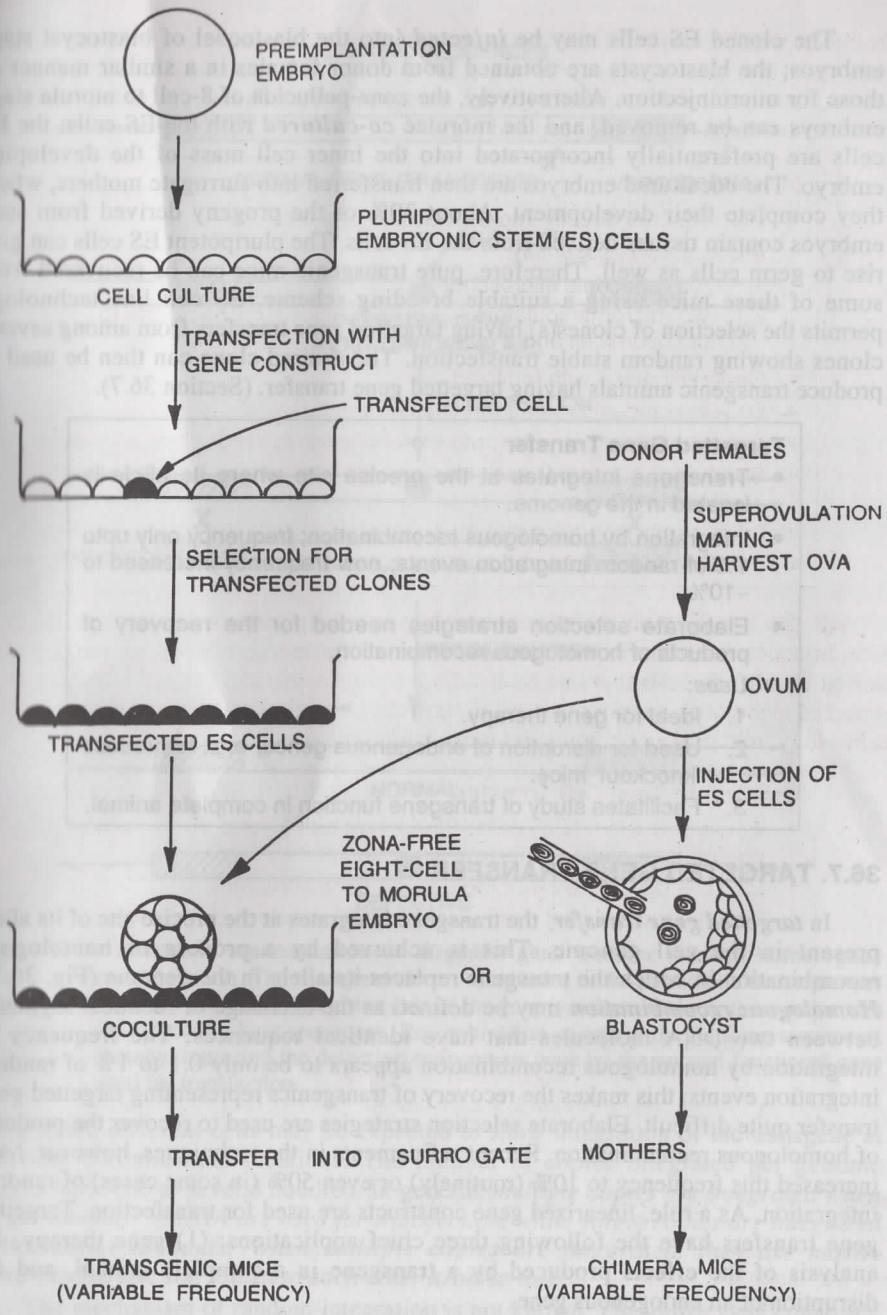


Fig. 36.6. A summary of the steps in embryonic stem (ES) cell transfer technology. When ES cells are injected into blastocoel embryos, chimaeric mice are obtained. But when 8-cell to morula stage embryos are co-cultured with ES cells, nonchimaeric transgenic mice are recovered.

The cloned ES cells may be ***injected*** into the blastocoel of blastocyst stage embryos; the blastocysts are obtained from donor females in a similar manner as those for microinjection. Alternatively, the zona-pellucida of 8-cell to morula stage embroys can be removed, and the morulae ***co-cultured*** with the ES cells; the ES cells are preferentially incorporated into the inner cell mass of the developing embryo. The cocultured embryos are then transferred into surrogate mothers, where they complete their development. About 30% of the progeny derived from such embryos contain tissues derived from the ES cells. The pluripotent ES cells can give rise to germ cells as well. Therefore, pure transgenic mice can be recovered from some of these mice using a suitable breeding scheme. ES cell line technology permits the selection of clones(s) having targetted gene transfers from among several clones showing random stable transfection. The desired clone can then be used to produce transgenic animals having targetted gene transfer. (Section 36.7).

Targetted Gene Transfer

- Transgene integrates at the precise site where its allele is located in the genome.
- Integration by homologous recombination; frequency only upto 1% of random integration events; now frequency increased to ~10%.
- Elaborate selection strategies needed for the recovery of products of homologous recombination.

Uses:

1. Ideal for gene therapy.
2. Used for disruption of endogenous genes, *e.g.*, to produce 'knockout' mice.
3. Facilitates study of transgene function in complete animal.

36.7. TARGETED GENE TRANSFER

In ***targetted gene transfer***, the transgene integrates at the precise site of its allele present in the cell genome. This is achieved by a process of homologous recombination in which the transgene replaces its allele in the genome (Fig. 36.7). ***Homologous recombination*** may be defined as the exchange of identical segments between two DNA molecules that have identical sequences. The frequency of integration by homologous recombination appears to be only 0.1 to 1% of random integration events; this makes the recovery of transgenics representing targetted gene transfer quite difficult. Elaborate selection strategies are used to recover the products of homologous recombination. Recent refinements in the techniques, however, have increased this frequency to 10% (routinely) or even 50% (in some cases) of random integration. As a rule, linearized gene constructs are used for transfection. Targetted gene transfers have the following three chief applications: (1) gene therapy, (2) analysis of the effects produced by a transgene in a complete animal, and (3) disruption of an endogenous gene.

36.8. TRANSGENE INTEGRATION

In general, *transgenes integrate at random sites in any of the chromosomes of the genome of host cells. Usually, in a given cell, integration occurs at a single location.*

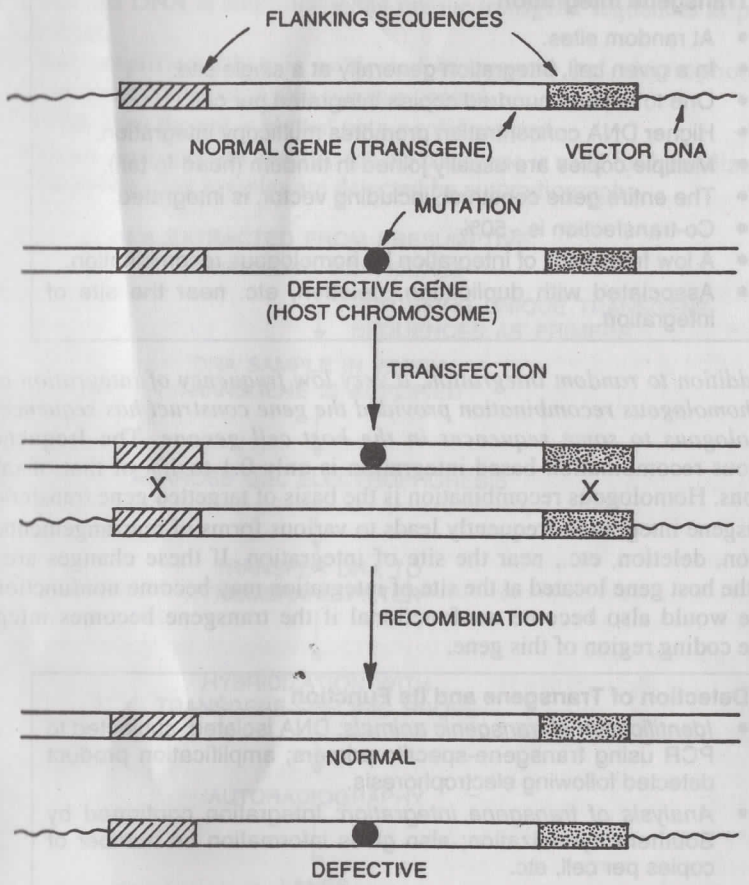


Fig. 36.7 A schematic representation of targetted gene transfer through homologous recombination. The transgene construct includes on its either side the sequences naturally present on the either side of the gene it aims to replace; these sequences are called flanking sequences. Recombination occurs in the flanking sequences, thereby, replacing the defective endogenous gene by the normal functional gene used for transfection.

As a result, different cells may be expected to show integration of the transgene at different chromosomal locations. The number of copies integrated per genome ranges from one to several hundred. In general, *multiple copies are integrated when larger amounts of DNA are used for transfection, while single copies are integrated with smaller amounts. When multiple copies are integrated, they are mostly integrated at one site joined to each other head-to-tail.*

The mechanism of random integration is not known. *The entire gene construct, including the vector DNA, becomes integrated. When two different gene constructs are mixed and used for transfection, they tend to be integrated together at the same site; this is known as* ***co-transfection.***

Transgene Integration

- At random sites.
- In a given cell, integration generally at a single site.
- One to several hundred copies integrated per cell.
- Higher DNA concentration promotes multicopy integration.
- Multiple copies are usually joined in tandem (head-to tail).
- The entire gene construct, including vector, is integrated.
- Co-transfection is ~50%.
- A low frequency of integration by homologous recombination.
- Associated with duplication, deletion, etc. near the site of integration.

In addition to random integration, a very low frequency of integration occurs through homologous recombination provided the gene construct has sequences that are homologous to some sequences in the host cell genone. The frequency of homologous recombination-based integration is only 0.1 to 1% of that of random integrations. Homologous recombination is the basis of targetted gene transfers.

Transgene integration frequently leads to various forms of rearrangements, *e.g.*, duplication, deletion, etc., near the site of integration. If these changes are large enough, the host gene located at the site of integration may become nonfunctional. A host gene would also become nonfunctional if the transgene becomes integrated within the coding region of this gene.

Detection of Transgene and Its Function

- *Identification of transgenic animals*: DNA isolated; subjected to PCR using transgene-specific primers; amplification product detected following electrophoresis.
- *Analysis of transgene integration*: Integration confirmed by Southern hybridization; also gives information on number of copies per cell, etc.
- *Detection of mRNA production*: Purified RNA subjected to RNA dot blot hybridization or northern hybridization.
- *Detection of protein production*: Proteins are detected using (1) antibodies, or (2) on the basis of enzymatic activity of the transgenic protein.

36.9. DETECTION OF TRANSGENES

It is essential to identify the transgenic animals and to establish that they have stably integrated transgenes, which are transcribed and the protein encoded by the transgene is produced. These may be achieved as follows.

36.9.1. Identification of Transgenic Animals

Transgenic individuals are generally identified by either dot-blot technique or PCR amplification of the transgenes. Genomic DNAs are extracted from tail biopsies from 6-7 week-old mice or from pectoral fin tissues of fishes. The PCR approach is briefly described below (Fig. 36.6).

1. The test DNA is amplified using unique transgene sequences as primers in a PCR.
2. The amplified DNA is subjected to agarose gel electrophoresis; the transgene construct is also run as a control.
3. DNA from the gel is blotted onto a solid support.
4. A radioactive probe specific for the transgene is used for hybridization, and the hybridizing samples are detected by autoradiography.

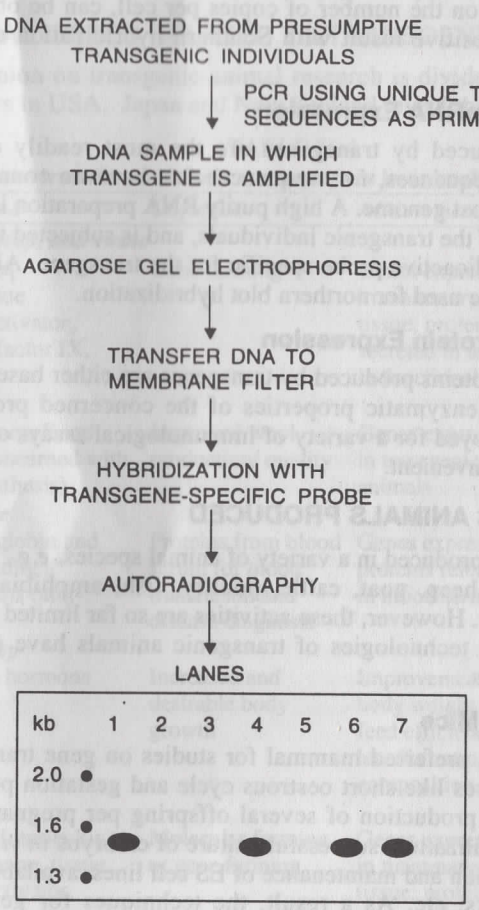


Fig. 36.8. A schematic representation of the use of PCR for identification of transgenic animals. Kb, kilobase (size is determined by running in the gel DNAs of known size); Lane 1, transgene construct used for transfection (control); Lanes 2-7, PCR amplified DNAs from presumptive transgenic animals (lanes 4, 6. 7 test positive for the transgene).

The samples that test positive for hybridization with the probe are from suspected transgenic individuals. However, both dot blot and PCR techniques detect the transgene irrespective of whether it is integrated in the genome or it is present in an extra chromosomal state.

36.9.2. Analysis of Transgene Integration

The integration of transgene into the genome is confirmed by Southern hybridization of genomic DNA extracted from the suspected transgenic individuals. The DNA is digested with a suitable restriction enzyme, and then subjected to electrophoresis. Radioactive probes specific for the transgenes are used for hybrization, after the DNA present in the gel is denatured and blotted onto a solid support. Using this approach, integration of transgene can be established beyond doubt, and information on the number of copies per cell, can be obtained. All the individuals that give positive result with Southern hybridization are regarded as confirmed transgenics.

36.9.3. Detection of mRNA Expression

The mRNAs produced by transgenes are the most readily detected if the mRNAs have unique sequences, *i.e.*, sequences that have no counterparts among those produced by the host genome. A high purity RNA preparation is obtained from the appropriate tissue of the transgenic individuals, and is subjected to RNA dot blot hybridization with a radioactive probe specific for the transgene. Alternatively, the RNA preparation may be used for northern blot hybridization.

36.9.4. Assays for Protein Expression

Detection of the proteins produced by transgenes are either based on antibodies specific for or on the enzymatic properties of the concerned proteins. Specific antibodies can be employed for a variety of immunological assays of which ELISA is the most rapid and convenient.

36.10. TRANSGENIC ANIMALS PRODUCED

Transgenics have been produced in a variety of animal species, *e.g.*, mice (including rats), rabbits, swine, sheep, goat, cattle, poultry, fish, amphibians, insects and nematodes (Table 36.4). However, these activities are so far limited to experimental stages, and production technologies of transgenic animals have not yet reached commercial applications.

36.10.1. Transgenic Mice

Mouse is the most preferred mammal for studies on gene transfers due to its many favourable features like short oestrous cycle and gestation period, relatively short generation time, production of several offspring per pregnancy (*i.e.*, litter), convenient *in vitro* fertilization, successful culture of embryos *in vitro* at least for a period of time, production and maintenance of ES cell lines, availability of a diverse array of genetic stocks, etc. As a result, the techniques for gene transfer and transgenic production have been developed using mice as model; subsequently, these approaches have been modified to adapt them to other animal species. More recently, rats and rabbits are being increasingly used for research work on gene transfer.

36.10.2. Transgenic Cattle

The only successful transfection technique in cattle is microinjection of fertilized ova, which may either be recovered surgically or may be obtained from ovaries extracted from slaughtered cows and cultured *in vitro*. The two chief objectives of transgenic production are as follows: (1) increased milk or meat

production and (2) molecular farming. Several human genes have been successfully transferred in cows and expressed in the mammary tissue. The transgene encoded protein is secreted in milk from where it is easily harvested (Table 36.4). Other objectives for transgenic cattle production are improved properties and proportions of caseins, contents of lactose and butterfat in milk, and enhanced resistance to viral and bacterial pathogens, including the development of constitutive immunity due to germ line transmission of rearranged antibody genes specific to selected pathogens.

36.11. ETHICAL ISSUES RELATED TO TRANSGENIC ANIMALS

The social opinion on transgenic animal research is divided almost in the middle. Opinion surveys in USA, Japan and New Zealand reveal that only 42, 54 and 58%,

TABLE 36.4

Some notable gene transfers and their consequences in animals of economic importance

Gene transferred	*Objective*	*Achievements*	*Remarks*
Cattle, goat, sheep, and swine			
Human genes α_1 antitrypsin, tissue plasminogen activator, blood clotting factor IX, and protein C	Gene farming	Genes expressed in mammary tissue; proteins secreted in milk in functional form	Experimental stages
Sheep			
Bacterial genes *cysE* and *cysM* (genes concerned with cysteine biosynthesis)	Improved wool production/ quality	Genes expressed in transgenic animals	—
Mice and swine			
Human haemoglobin and specific circulating immunoglobulins (antibodies)	Proteins from blood serum for blood transfusion and disease diagnosis	Genes expressed; proteins released in blood serum	Experimental stages
Swine and sheep			
Human growth hormone	Increased and desirable body growth	Improvements in body weight gain, feed efficiency, lean/fat ratio, fat composition, etc.	Produces joint pathology, skeletal defects, ulcers, infertility, etc.
Rabbits			
Human genes: Interleukin 2, growth hormone, tissue plasminogen activator, α_1- antitrypsin, etc.	Molecular farming or gene farming	Genes expressed in mammary tissue; proteins harvested from milk	Experimental stages
Bovine gene: *a*-lactoglobulin			
Fish			
Salmon or rainbow trout growth hormone	Increased body growth	Upto 60% increase in size	Transgenes stably inherited; growth improved by selection
Anti-freeze protein gene, α-globin gene, chicken δ-crystalline protein gene, etc.	Variable	Genes expressed in transgenic individuals	Transgenes are stably inherited

respectively, of the people participating in the survey favour such research. The main reasons for opposition of people may be summarised as follows.

1. Use of animals in biotechnological research causes great suffering to the animals.
2. It is felt that by using animals for the production of pharmaceutical proteins we reduce them to mere factories.
3. Some people feel that animals should be regarded as equal to humans in that they have the same basic rights as human beings.
4. An argument attempts to focus on the integrity of species in that each biological species has a right to exist as a separate identifiable entity. Gene transfers are considered to voilate this integrity.
5. Finally, the introduction of human genes into animals, and *vice-versa,* may be seen by many as clouding the definition of "humanness".

36.12. CONCLUSIONS AND PERSPECTIVES

The non-availability of ES cell technology for mammals other than mice has prevented targetted gene transfer in them. ES cell or ES cell-like lines have been developed in hamster, rat, rabbit and pig, but so far transgenic animals have not been produced using this technology. Targetted gene transfer has enabled the production of "knockout" mice which have been extremely valuable in studies on mouse genetics and molecular biology. Therefore, concerted efforts need to be made to develop the ES cell technology for other animals as well.

Transgenic animals have been very useful in biological, biomedical and biotechnological researches, and they offer unique opportunities in animal husbandry, dairy and pharmaceutical industries. The potential is immense, and a clear indication is available from the results obtained so far. However, none of the approaches has so far reached the commercial stage, which signifies the critical gap between promise and fulfilment. Future efforts no doubt will focus on this gap, and are expected to generate valuable animal lines/ products to enhance human welfare and enrich the quality of human life.

QUESTIONS

1. What are transgenic animals ? Describe the production of transgenic animals under the following heads: (*i*) objectives of gene transfer, (*ii*) vector, (*iii*) promoters, (*iv*) reporter genes and (*v*) methods of transfection.
2. Briefly describe the various transfection technique for animal cells and ova, and explain random and targetted gene integrations.
3. What are embryonic stem cells ? Describe the embryonic stem cell transfer technology for the production of transgenic mice and discuss its various applications and limitations.
4. Describe briefly the techniques for identification/selection of transgenic individuals, confirmation of transgene integration, transcription and protein production.
5. Briefly describe some cases of important gene transfers in some animal species. Explain the various objectives for gene transfers, and outline the ethical objections to the production of transgenic animals.

6. Write short notes on the following:
   (*i*) ethical issues related to transgenic animal production,
   (*ii*) targetted gene transfer,
   (*iii*) embryonic stem cell transfer technology,
   (*iv*) transgene integration,
   (*v*) detection of transgene integration and function,
   (*vi*) applications of transgenic animals,
   (*vii*) promoters used for gene expression,
   (*viii*) selectable reporter genes,
   (*ix*) transfection,
   (*x*) lipofection,
   (*xi*) calcium phosphate precipitation,
   (*xii*) microinjection,
   (*xiii*) electroporation.

37

Transgenic Plants

37.1. INTRODUCTION

The most potent biotechnological approach for gene transfer is the transfer of specifically constructed gene assemblies through various genetic transformation techniques; this constitutes ***genetic engineering.*** The plants obtained through genetic engineering contain a gene or genes usually from an unrelated organism; such genes are called ***transgenes,*** and the plants containing transgenes are known as ***transgenic plants.***

Transgenic Plants

- Plants that have a gene or genes transferred by genetic engineering.
- *Applications:*
  1. Studies on gene function.
  2. Gene isolation.
  3. Resistance to biotic stresses, *e.g.*, insects, pathogens, etc.
  4. Resistance to abiotic stresses, *e.g.*, drought, cold, etc.
  5. Improved quality.
  6. Production of novel biochemicals.
  7. Vaccine production, *e.g.*, edible vaccines.

37.2. APPLICATIONS OF TRANSGENIC PLANTS

Transgenic plants have both basic and applied uses, which are briefly summarised below.

1. They have proved to be extremely valuable tools in studies on plant molecular biology, regulation of gene action, identification of regulatory/promotary sequences, etc.
2. T-DNA and transposable elements produce mutations by becoming inserted within genes. Thereby, they act as 'molecular tags' for gene identification and isolation.
3. Specific genes have been transferred into plants to improve their resistance to abiotic stresses, *e.g.*, herbicides, drought, cold, etc.
4. Gene for resistance to various biotic stresses have been engineered to generate transgenic plants resistant to insects, viruses, etc.
5. Several gene transfers have been aimed at improving the produce quality.

6. Transgenic plants are aimed to produce novel biochemicals like hirudin, etc., or useful biopolymers like polyhydroxybutyrate, which are not produced by normal plants.
7. Transgenic plants can be used as edible vaccines.

37.3. GENE CONSTRUCTS

Gene construct signifies the assembly of various DNA sequences designed for easy identification of the transgene construct and for its efficient expression in transgenic individuals. In addition to the gene and its regulatory sequences, gene constructs usually have reporter genes for an easy identification and or selection of the transgenic individuals.

A typical plant gene has the following regions beginning with the 5'-end: (*i*) promoter (for transcription initiation), (*ii*) enhancer/silencer (concerned with regulation), (*iii* transcriptional start or cap site, (*iv*) an untranslated region or leader sequence, (*v*) initiation codon, (*vi*) exons, (*vii*) introns, (*viii*) the stop codon(s), (*ix*) a second untranslated region and, finally, (*x*) a poly(A) tail sequence, which is typically G/AATAA(A)*n* (Fig. 37.1).

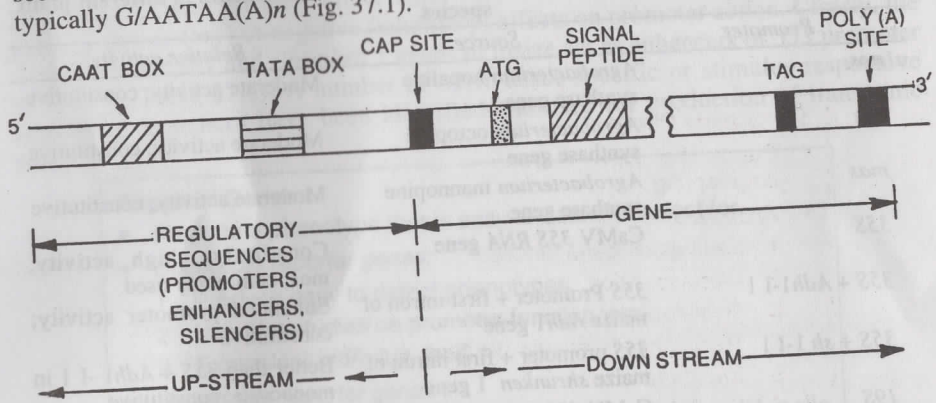


Fig. 37.1. Organization of a typical plant gene. CAAT box and TATA box, consensus sequences within promoters; cap site, transcription start site; ATG, (AUG in mRNA), translation initiation codon; TAG (UAG in mRNA), translation termination codon; poly(A) site, transcription termination site and polyadenylation site. Note that gene itself consists of exons (expressed sequences) and introns (interrupting sequences), which are not shown, and that the gene is depicted as broken since it is drawn much smaller than its normal size. The 5' and 3' indicate the direction of mRNA, which is identical in base sequence to that of the 'sense strand' of the gene; mRNA *itself is transcribed from the antisense strand.* The sequences beyond the cap site at the 5'-side are called ***upstream*** and the base positions from the startpoint are denoted by a – sign, while those on the 3'-side are called ***downstream*** and are designated by a + sign.

37.3.1. Promoters/Enhancers

A variety of promoter sequences have been used to drive genes in plant cells (Table 37.1). These include the *nos* (nonpaline syntheses), *ocs* (octopine synthase) and *mas* (mannopine synthase) promoters from *Agrobacterium,* and 35*S* RNA gene promoter of cauliflower mosaic virus (CaMV). The 35*S* promoter is the most commonly used constitutive promoter both in dicots and monocots; it is 10-40-fold

Promoters

- Binding site for RNA polymerase.
- Essential for transcription.
- *May be constitutive, i.e.,* gene expression in all cells at all the times.
- *Some are tissue specific*; gene expression in specific tissues only.
- *Some are stimulus-responsive, i.e.,* genes are expressed in response to specific stimuli.
- *Some may be development-specific.*
- *Enhancers* : Increase promoter function; involved in tissue-specific, etc. expression.
- *Silencers*: Suppress promoter function, usually in response to specific stimuli.

TABLE 37.1
Some of the promoters used for driving the expression of transgenes in different plant species

Promoter	*Source*	*Relative activity*
nos	*Agrobacterium* nopaline synthase gene	Moderate activity; constitutive
ocs	*Agrobacterium* octopine synthase gene	Moderate activity; constitutive
mas	*Agrobacterium* mannopine synthase gene	Moderate activity; constitutive
35S	CaMV *35S RNA* gene	Constitutive, high activity; most commonly used
35S + *Adh*1-I 1	*35S* Promoter + first intron of maize *Adh*1 gene	Enhanced promoter activity; constitutive
35S + *sh* 1-I 1	*35S* promoter + first intron of maize *shrunken* 1 gene	Better than *35S* + *Adh*1 -I 1 in monocosts; constitutive
19S	CaMV *19S RNA* gene	Constitutive; moderate activity
Gene VI	CaMV *gene VI* encoding the matrix protein	Constitutive expression
*Adh*1	Promoter of alcohol dehydrogenase gene of maize	Moderate activity in cereals; anaerobic expression
Emu	Modified from *Adh*1 promoter and its first intron	Moderate activity in cereals; anaerobic expression
Act 1 + *Act*1-I 1	Rice *actin* gene + its first intron	Moderate activity; constitutive
*Ubi*1 + *Ubi*1-I 1 (or I 6)	Maize *ubiquitin* gene promoter + its first (or sixth) intron	High activity in cereals; constitutive
PHA-L	*Phytohaemagglutinin* gene of *P. vulgaris*	Strong, seed- specific activity
Vicilin promoter	Pea *vicilin* storage protein gene	Seed-specific promoter
Extensin promoter	Carrot *extensin* gene long transcript promoter	—

more efficient than the *nos* promoter. The maize alcohol dehydrogenase 1 (*Adh*1) promoter shows anaerobic induction, *i.e.,* expression in roots. The *Adh*1 promoter activity in monocots is either comparable to or higher than that of *35S* promoter, which is not an efficient promoter for monocot species. The *Adh*1 intron 1 and maize *shrunken* 1 locus intron 1, when placed between the promoter and the gene, enhance the efficiency of CaMV *35S* promoter, the *shrunken* gene intron 1 being, in some cases, upto 10-fold more effective than the *Adh*1 intron.

Expression of many genes is confined to specific tissues and/or is induced by specific stimuli; such genes are called ***tissue-specific*** or ***stimulus-responsive genes,*** respectively. Such specificities in gene expression are due to certain DNA sequences called enhancers/silencers, either lying within or at a considerable distance (upto several kilobases) away from the promoters they affect. An ***enhancer*** may be defined as a DNA sequence, which increases the activity of promoter of a gene, while DNA sequences suppressing promoter activity are called ***silencers.*** Thus enhancer/ silencer sequences regulate the activity of promoters but are not themselves involved in the promoter activity *per se.* These sequences act on promoters located on their either side, vary considerably in size, and can exert positive (promotory) or negative (inhibitory) effects on promoter action. Clearly, the first intros of *Adh*1 and *shrunken* 1 genes of maize act as enhancers of *35S* promoter of CaMV (Table 37.1). A number of other tissue-specific or stimulus responsive regulating sequences have been identified and used for production of transgenic plants.

Reporter Genes

- Produce a phenotype that is easy to detect of select for.
- *Socarable reporter genes:*
  1. Produce easy to detect phenotypes.
  2. Used for studies on promoter function, etc.
  3. Examples: *uidA, lux, lacZ.*
- *Selectable reporter genes* :
  1. Produce a phenotype that allows selection of cells expressing the gene.
  2. Used for production of transgenic plants.
  3. Linked to the transgene construct being transferred.
  4. *Examples*: genes for herbicide (*bar, deh*) or antibiotic (*npt*II, *hptIV, kan*^{r}) resistance.

37.3.2. Reporter Genes

The expression of some bacterial (usually), animal or plant genes is easily detected either through highly sensitive enzyme assays (***scorable reporter genes***) or through expression of resistance to a toxin (***selectable reporter genes***); such genes are called ***reporter genes***. Some commonly used easily detectable enzyme producing genes are *nos* (nopaline synthase, from *Agrobacterium*), *lux* (luciferase from bacteria or firefly), *cat* (chloramphenicol acetyltransferase from bacteria), and *uidA* (β-glucuronidase from bacteria), etc. (Table 37.2).

The essential features of an ideal reporter gene are (1) lack of endogenous activity in plant cells of the concerned enzyme, (2) an efficient and easy detection,

TABLE. 37.2
Some of the marker genes used for gene transfer into plants

Gene	Source	Characteristic
	Scorable Reporter Genes	
ocs (octopine synthase)	*Agrobacterium*	Octopine production
nos (napoline synthase)	*Agrobacterium*	Nopaline production
cat (chloramphemicol acetyltransferase)	*E. coli*	Acetylation of chloramphenicol
lux (luciferase)	Bacteria, firefly	Phosphorescence
lacZ (β-galactosidase)	*E. coli*	Blue colour (substrate X-gal)
uidA (β-glucuronidase)	*E. coli*	Blue colour (substrate X-gluc).
Green fluorescent protein	Jelly fish	Green fluorescence
Diverse regulatory factors involved in anthocyanin biosynthesis	Maize	Enhanced anthocyanin pigmentation
	Selectable Reporter Genes	
dhfr (dihydrofolate reductase)	Tn7, mouse, plasmid R67	Methotrexate resistance
nptII (neomycin phosphotransferase)	Tn5	Kanamycin, neomycin and G418 resistance
hptIV (hygromycin phosphotransferase)	*E. coli, Streptomyces hygroscopicus*	Hygromycin B resistance
bar (phosphinothricin acetyl-transferase)	*Streptomyces hygroscopicus*	Phosphinothricin resistance
*deh*1 (Dehalogenase)	*Pseudomonas putida*	Dalapon resistance
— (Streptomycin and spectinomycin phosphotransferase)	*E. coli*	Resistance to streptomycin and spectinomycin
ALS (Mutant acetolactate synthase)	*Arabidopsis thaliana,* tobacco (*N. tabacum*)	Resistance to sulfonyl ureas
Kan^r [Aminoglycoside 3'-phosphotransferase II; APH (3)II]	Bacteria	Resistance to kanamycin

and (3) a relatively rapid degradation of the enzyme. Each reporter gene has some advantages and some disadvantages and none of them is ideally suited for all plant species. For example, there is little or no endogenous lucifearse activity in plant cells, the enzyme serves as a visual marker, the assay is quite sensitive, but the enzyme is highly stable. Similarly, the GUS enzyme serves as visual marker and the assay is highly sensitive so that expression in individual cells can be monitored, but the enzymes is highly stable and many plant species show variable endogenous GUS activity. *Reporter genes, more particularly, scorable reporter genes, have been extensively used to assay the function of promoters and other regulatory sequences, and also to demonstrate the transformation of plant cells.*

A key step in gene transfer is the selection of few transformed cells from among the nontransformed ones. This is usually achieved by the use of ***selectable markers,*** which are genes that confer resistance to various compounds toxic to normal plant cells. The best selection agents are those that either inhibit growth or slowly kill the nontransformed cells so that the dying cells do not overwhelm the transformed ones. The commonly used selectable marker genes include those conferring resistance to

methotrexate (*dhfr,* isolated from Tn7 or plasmid R67), the antibiotics kanamycin (*npt*II, encoding neomycin phosphotransferase), hygromycin (*hptIV*, encoding hygromycin phosphotransferase, isolated from *E. coli*), broad range herbicides glyphosate (modified versions of the enzyme EPSPS, 5-enolpyruvate shikimate- 3-phosphate synthase, isolated from *E. coli* or *Salmonella typhimurium*), phosphinothricin (*bar*, isolated from *Streptomyces hygroscopicus,* codes for phophinothricin acetyltransferase), dalapon (*deh*1, isolated from *Pseudomonas putida,* codes or dehalogenase), etc. (Table 37.2). Each selectable marker presents some favourable and some unfavourable features. Therefore, the choice of marker should be based on the plant species and other considerations in the study. The marker gene of choice is fused with a suitable plant promoter and is placed in a suitable vector along with the gene being transferred.

Vectors

- A *vector* is a DNA molecule that functions as replicon, and serves as a vehicle for the DNA fragment being transferred.
- *Types*: Two types.
  1. *Plasmid vectors:* Based on Ti plasmid of *Agrobacterium tumefaciens* or Ri plasmid of *A. rhizogenes.* Used for stable transformation.
  2. *Virus vectors*: Derived from caulimo and gemini viruses; used for production of multiple copies of transgene or large quantities of the transgene-encoded protein.

37.4. VECTORS

Plant cells, like animal cells, also take up naked DNA fragments, *i.e.,* DNA molecules not integrated into any vector, added to the medium; this DNA becomes integrated into the plant genome. However, the efficiency of transformation is very low. Therefore, DNA segments are usually integrated into a vector.

A ***vector*** is usually a circular DNA molecule capable of independent existence and replication, *e.g.,* plasmids, viruses, etc. In case of plants, *Agrobacterium* plasmids Ti and Ri are the most commonly used vectors. Both the plasmids have a T-region, which is transferred into plant cells and is integrated into their genomes. The genes to be transferred are placed within the T-DNA of these plasmids.

In addition, DNA viruses like caulimo viruses and gemini viruses can be used as vectors for gene amplification since viruses multiply in plant cells. But the virus genomes are neither integrated into plant genomes nor they are generally transmitted through seed. Hence these vectors cannot be used for stable gene transfers. The cauliflower mosaic virus (CaMV) has been extensively studied but has so far found only limited use as a vector; gemini viruses may be more promising. Intact or modified plant viruses can be readily introduced into plant cells using *Agrobacterium* Ti or Ri plasmid as a vector, the viral genome being placed within their T-DNA; this phenomenon is called ***agroinfection.*** Agroinfection is feasible both in monocots and dicots, and *A. rhizogenes* seems to be more infective than *A. tumefaciens* in the case of cereals.

37.4.1. *Agrobacterium*

A. tumefaciens and *A. rhizogenes* are plant pathogenic, gram-negative soil bacteria known for over 90 years to cause crowngall (*A. tumefaciens*) and hairy

A. tumefaciens
- Gram-negative soil bacterium.
- Produces crowngall disease in dicots.
- Infects plant cells near wounds.
- Crowngall produced due to transfer of T-DNA from Ti plasmid into the plant genome.
- Uses the opines produced by crowngall cells as its carbon and nitrogen source.

root diseases, respectively, of dicot plants. They infect plant cells near wounds, usually at the crown of roots at the soil surface. The process of infection by *A. tumefaciens* culminates in the transfer of a small part of *p*Ti into the plant cell genome; this DNA sequence is celled ***T-DNA*** (***transferred DNA***) (Fig. 37.2). The infection process is governed by both chromosomal and the plasmid-borne genes of *A. tumefaciens*. Infection begins when *Agrobacterium* cells become attached to plant cells; this step is governed by bacterial chromosomal genes, *e.g.*, *chv* (chromosomal virulence) genes. These chromosomal genes are expressed constitutively, *i.e.*, expressed in all bacterial cells at all the times.

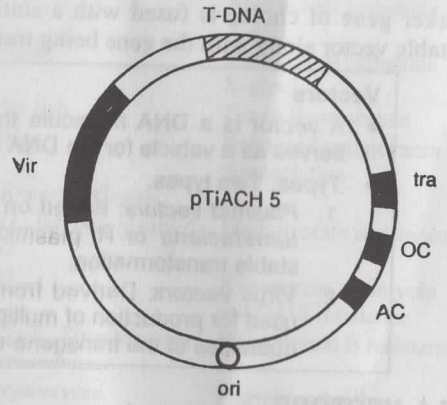


Fig. 37.2. The general organisation of octopine type Ti plasmid. *Vir*, virulence region; *tra*, conjugative transfer, OC, octopine catabolism; AC, agropine catabolism; *ori*, origin of replication in *Agrobacterium*.

Crowngall Cells
- Grow on GR-free medium.
- Produce tumours when grafted onto healthy plants.
- Produce opines, which are unique nitrogenous compounds not produced by normal plant cells.
- Opine production occurs even when the cells are cultured *in vitro* or plants are regenerated from them.
- Type of opine produced depends on *Agrobacterium* strain that produced the crowngall.
- Their genome contains the T-DNA from Ti plasmid.

37.4.2. Properties of Crowngall Cells

Infection by *A. tumefaciens* produces tumour like growth from which roots and/or shoots may sometimes be produced. But infection by *A. rhizogenes* gives rise to 'hairy roots'; in some species, shoots may regenerate from the roots giving rise to complete plants.

1. Both hairy root and crowngall cells (freed from *Agrobacterium* cells) are capable of growing in culture on a growth regulator (GR)-free medium.
2. They generally induce tumours when grafted onto healthy plants.
3. The crowngall and hairy root cells synthesize unique nitrogenous compounds celled ***opines***, which are neither produced by normal plant cells nor utilized by them. *Agrobacterium* cells use opines as their carbon and nitrogen source. The crowngall cells continue to produce opines even when they are cultured *in vitro*; even the plants regenerated from these cells produce them.
4. The type of opine produced by a crowngall depends on the bacterial strain (more specifically on the *Agrobacterium* plasmid) producing the crowngall.
5. A bacterial strain produces only one type of opine, and it also catabolizes only that opine; the concerned genes are present in the T- DNA of *p*Ti or *p*Ri.

Ti Plasmid
- Tumour (crowngall) inducing plasmid.
- A ~200 kb megaplasmid found in *A. tumefaciens.*
- Two regions of Ti plasmid critical in gene transfer:
  1. *T- DNA* (transferred *DNA*).
  2. *vir* (*vir*lence) region.
- *Virulence* is the ability of the bacterium to produce crowngall.

37.4.3. The Ti Plasmid

The Ti plasmid is a large conjugative plasmid of about 200 kb. *p*Ti is lost when *Agrobacterium* is grown above 28°C. (1) *p*Ti contains some genes (the genes located within their T-DNA), which have regulatory sequences recognised by plant cells, while their remaining genes have prokaryotic regulatory sequences. (2) *p*Ti naturally transfers a part of its DNA, the T-DNA, into the host plant genome, which makes *Agrobacterium* a natural genetic engineer.

T-DNA
- ~ 23 kb segment.
- Ends in a 24 bp direct repeat at the right and left borders.
- Contains following genes:
  1. *iaaM* and *iaaH* (auxin biosynthesis).
  2. *ipt* (cytokinin biosynthesis).
     (These together are called *ancogenes*)
  3. genes for opine biosynthesis and a few others.
- All genes have eukaryotic regulatory sequences and are expressed only in plant cells.
- The two borders are essential for T-DNA transfer.

37.4.4. Organisation of T-DNA

T-DNA (***transferred DNA***) is ~23 kb segment of Ti plasmid, which is transferred into the plant genome during *Agrobacterium* infection. T-DNA is defined on both the sides by a 24 bp direct repeat border sequence, and contains the genes for tumour induction and those for opine biosynthesis. *p*Ti T-DNA has 3

genes, which are involved in crowngall formation. Two of these genes (*iaaM* and *iaaH*) encode enzymes, which together convert tryptophan into IAA (indole-3-acetic acid). The third gene, *ipt*, encodes an enzyme, which produces the zeatin-type cytokinin isopentenyl adenine (Fig. 37.3).

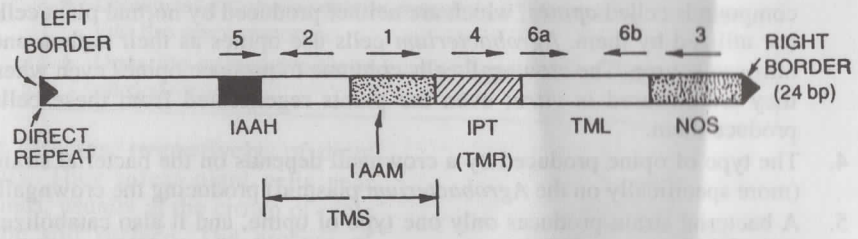


Fig. 37.3. Organisation of a nopaline type Ti plasmid T-DNA. Arrows indicate the direction of transcription and numbers indicate the transcriptional units. *iaaM*, encodes tryptophan-2-monooxygesase; *iaaH*, encodes indoleacetamide hydrolase [both involved in auxin biosynthesis, together constitute the *tms* (shooty tumour) locus]; *ipt*, encodes isopentenyl transferase (cytokinin biosynthesis); *tml*, large tumour locus; *nos*, encodes nopaline synthase.

All the genes present in T-DNA contain eukaryotic regulatory sequences. As a result, these genes are expressed only in plant cells, and they are not expressed in the Agrobacterium. *Further, only the 24 bp left and right borders of T-DNA are essential for T-DNA transfer.*

37.4.5. Organisation of *vir* Region

The *vir* region contains 8 operons (designated as *virA*, *virB*, *virC*, *virD*, *virE*, *virF*, *virG* and *virH*), which together have 25 genes (Fig. 37.4). This region mediates the transfer of T- DNA into plant genomes, and hence is essential for ***virulence*** or production of crowngall; therefore, it is called the ***virulence region*** or ***vir region***. The genes of *vir* region are **not transferred** themselves; **they only induce the transfer of T-DNA**. The operons *virA* and *virG* are constitutive, encode one protein each, and are concerned with the regulation of all the *vir* operons. The other *vir* operons encode various proteins involved in T-DNA transfer (Table. 37.3).

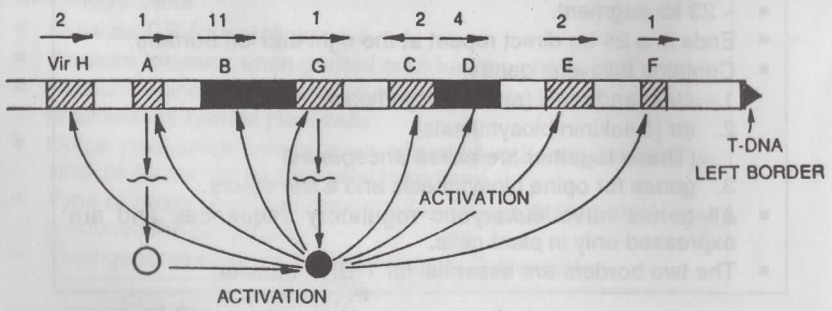


Fig. 37.4. Organisation of the *vir* region of an octopine type *p*TI. Operons *virA* and *virG* show low levels of constitutive expression. In response to plant phenolic signal molecules, VirA protein phosphorylates itself and VirG protein; VirG now induces all the 8 operons of *vir* regulon. The arrows, above the letters designating the operons, indicate the direction of transcription, and the numbers above the arrows represent the number of genes in different operons.

TABLE 37.3
The *vir* operons of *p*Ti (nopaline type) and their functions

Operon	*Function*
virA * (1)**	Encodes a sensor protein, receptor for acetosyringone and functions as an autokinase; also phosphorylates VirG protein; *constitutive expression.*
virB (11)	Membrane proteins; possibly form a channel for T-DNA transport (conjugal tube formation); VirB11 has ATPase activity.
virC (2)	Helicase; binds to the overdrive region just outside the right border; involved in unwinding of T-DNA.
virD (4)	VirD1 has topoisomerase activity; it binds to the right border of T-DNA; VirD2 is an endonuclease that nicks the right border.
virE (2)	Single strand binding proteins (SSBP); bind to T-DNA during its transfer.
virF (1)	Not well understood.
virG (1)	DNA binding protein; probably forms dimer after phosphorylation by VirA, and induces the expression of all *vir* operons (operons *A* to *H*); *constitutive expression.*
virH(2)	Not well known.

* *virA*, *virB*, etc. are the various operons of the *vir* region.
** Figures within parentheses are the numbers of genes in the operons.

37.4.6. Transfer of T-DNA

T-DNA transfer is brought about by the *vir* region, which is activated by the phenolic signal molecules acetosyringone and α-hydroxyacetosyringone (Fig. 37.5); these are produced by wounded tissues of virtually all dicot plant species, and constitute the ***wound response.*** The phenolic signal molecules bind to the *virA* gene product, the VirA protein. VirA is located in the inner membrane of *Agrobacterium.* The VirA becomes activated in response to the signal molecules. It now functions as an autokinase and phosphorylates itself. The phosphorylated VirA then phosphorylates VirG protein, which then dimerises. VirG protein is a DNA binding

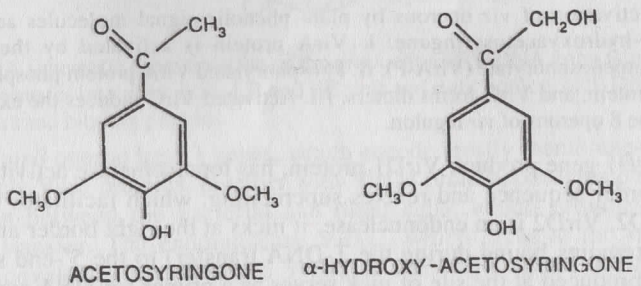


Fig. 37.5. Structures of acetosyringone and α-hydroxyacetosyringone. These signal molecules are produced by wounded plant tissues, and constitute the wound response. These compounds activate the *vir* operons whose function is essential for T-DNA transfer into plant cells.

Transfer of T-DNA

- Injured plant cells produce acetosyringone.
- Protein VirA is activated by autophosphorylation in response to acetosyringone.
- Activated VirA phosphorylates VirG protein; VirG now forms dimer.
- VirG dimer induces expression of all the 8 *vir* operons.
- VirD1 relaxes supercoiling at the right border of T-DNA.
- VirD2 nicks of right border; binds the 5'-end so produced.
- The 3′-end at the nick is used for DNA replication; this displaces the T-DNA single-strand.
- Nick at the left border produces T-DNA single strand.
- Single strand binding proteins bind and stabilize T-DNA single strand.
- VirB and VirD4 proteins form conjugal tube between bacterial and plant cells.
- VirD2 guides T-DNA single strand into plant nucleus.

protein and requires phosphorylation for its DNA binding function. Activated VirG protein induces the expression of all the 8 *vir* operons (Fig. 37.6).

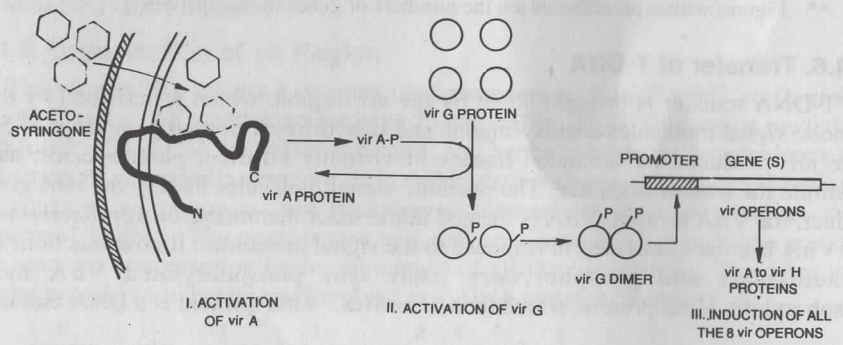


Fig. 37.6. Activation of *vir* operons by plant phenolic signal molecules acetosyringone/ α-hydroxyacetosyringone. I. VirA protein is activated by the phenolics to autophosphorylate (VirA-P). II. Phosphorylated VirA protein phosphorylates VirG protein, and VirG forms dimers. III. Activated VirG induces the expression of all the 8 operons of *vir* regulon.

The *virD*1 gene product, VirD1 protein, has topoisomerase activity. It binds to the right border sequence and relaxes supercoiling, which facilitates the action of protein VirD2. VirD2 is an endonuclease; it nicks at the right border and covalently binds (and remains bound during the T-DNA transfer) to the 5′-end so generated. The 3′-end produced at the site of nick serves as a primer for DNA synthesis in the 5′→ 3′ direction as a result of which the T-DNA strand is displaced from the DNA duplex (Fig. 37.7). The T-DNA strand is again nicked at the left border to generate a single-strand copy of T-DNA. VirE2 protein is a single-strand binding protein; about 600 molecules of it bind to the single-stranded T-DNA and protect it from nuclease action.

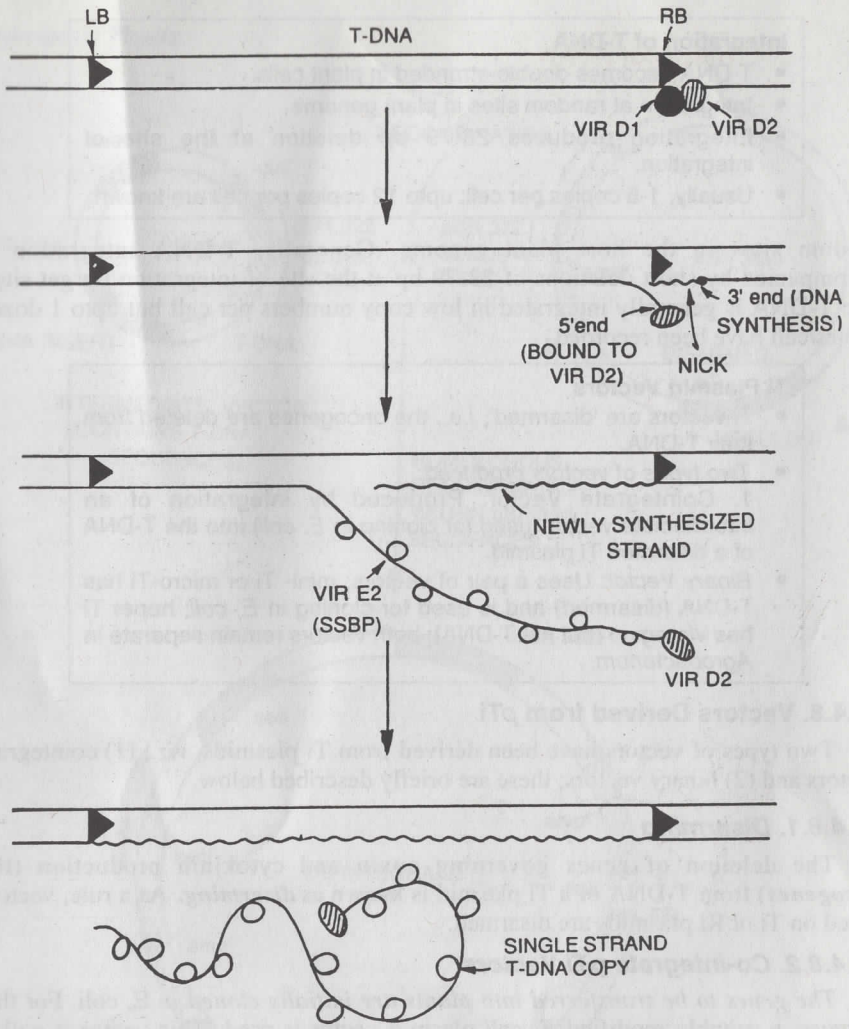


Fig. 37.7. A schematic representation of the production of T-DNA copy (single- stranded) for transfer into plant cells. LB and RB, left and right borders of T-DNA; SSBP, single-strand binding proteins.

The *virB* operon has 11 genes, which encode mostly membrane-bound proteins. Most likely VirB proteins, together with VirD4 protein, participate in conjugal tube formation between the bacterial and plant cells, which provides a channel for T-DNA transfer. The endonuclease VirD2, which nicks at the right border and remains covalently bound to the 5'-end of the single-strand T-DNA copy, has a signal sequence, which drives it into the nucleus of the transformed plant cell.

37.4.7. Integration of T-DNA into Plant Genome

T-DNA enters plant cells in a single stranded form, and is immediately converted into a double-stranded form. The double-stranded T-DNA integrates at

Integration of T-DNA

- T-DNA becomes double-stranded in plant cells.
- Integration at random sites in plant genome.
- Integration produces 23-79 bp deletion at the site of integration.
- Usually, 1-5 copies per cell; upto 12 copies per cell are known.

random sites in the host plant genome. Generally, T-DNA integration is accompanied by short deletions of 23-79 bp at the site of integration (target site). The T-DNA is generally integrated in low copy numbers per cell but upto 1 dozen copies/cell have been recorded.

Ti Plasmid Vectors

- Ti vectors are 'disarmed', *i.e.*, the oncogenes are deleted from their T-DNA.
- *Two types of vectors produced:*
  1. Cointegrate Vector: Produced by integration of an intermediate vector (used for cloning in *E. coli*) into the T-DNA of a disarmed Ti plasmid.
- *Binary Vector*: Uses a pair of vectors; mini- Ti or micro-Ti has T-DNA (disarmed) and is used for cloning in *E. coli*; heper Ti has *vir* region (but not T-DNA): both vectors remain separate in *Agrobacterium*.

37.4.8. Vectors Derived from *p*Ti

Two types of vectors have been derived from Ti plasmids, *viz.*, (1) cointegrate vectors and (2) binary vectors; these are briefly described below.

37.4.8.1. Disarming

The deletion of genes governing auxin and cytokinin production (the ***oncogenes***) from T-DNA of a Ti plasmid is known as ***disarming.*** As a rule, vectors based on Ti or Ri plasmids are disarmed.

37.4.8.2. Co-integrate p*Ti Vectors*

The genes to be transferred into plants are initially cloned in E. coli. For this purpose, a suitably modified *E. coli* plasmid vector is used. This vector is called ***intermediate vector***; it is subsequently integrated into a disarmed *p*Ti; this gives rise to the ***cointegrate vector***. A common approach is to disarm, a *p*Ti by replacing its oncogenes with sequences from the *E. coli* plasmid to be used for cloning the DNA insert/gene to be transferred (Fig. 37.8).

37.4.8.3. Binary Vectors

A ***binary vector*** consists of a pair of plasmids of which one plasmid contains disarmed T-DNA (at least the left and right borders of T-DNA must be present), while the other contains *vir* region, and ordinarily lacks the entire T-DNA including the borders. The plasmid containing disarmed T-DNA is called ***mini-Ti*** or ***micro-Ti***, *e.g.*, Bin 19, and has the origins for replication in both *E. coli* and *Agrobacterium*. The DNA insert in integrated within the T-region of mini-Ti and the recombinant DNA is cloned in *E. coli*. Recombinant *mini*-Ti is transferred into an *Agrobacterium* strain that has a 'helper' Ti plasmid (Fig. 37.9). This *Agrobacterium* is now used for

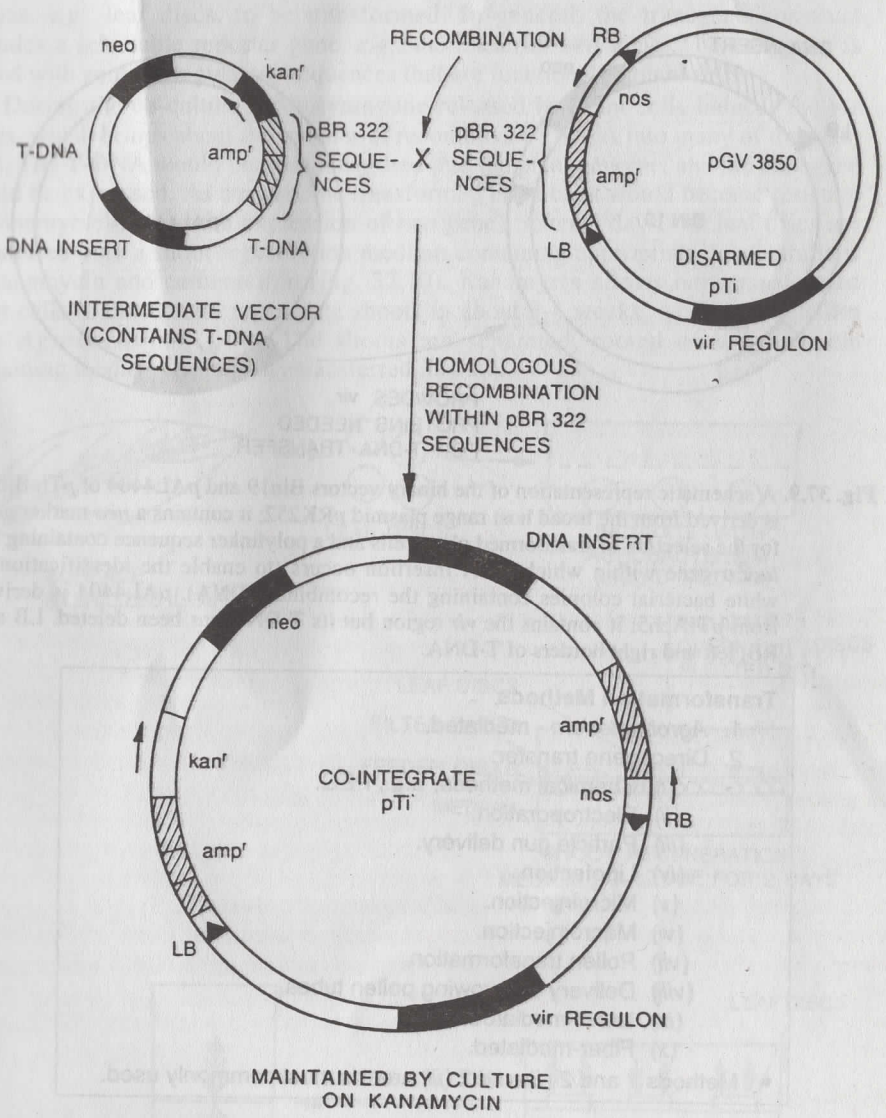


Fig. 37.8. Homologous recombination between a suitably disarmed *p*Ti and an IV (intermediate vector) containing the desired DNA insert to produce a co-integrate vector. Recombination occurs within the *p*BR322 sequences. LB and RB, left and right borders of T-DNA; *neo*, neomycin phosphotransferase; *kan*^{r} kanamycin resistance; *amp*^{r}, ampicillin resistance.

transformation of plant cells. A ***helper Ti*** plasmid has a functional *vir* region but lacks T-DNA. The *vir* genes of helper Ti induce the transfer of T-DNA present in the recombinant mini-Ti vector that is present in the same bacterial cell. As a result, the gene construct present within the T-region is also transferred into the plant cell.

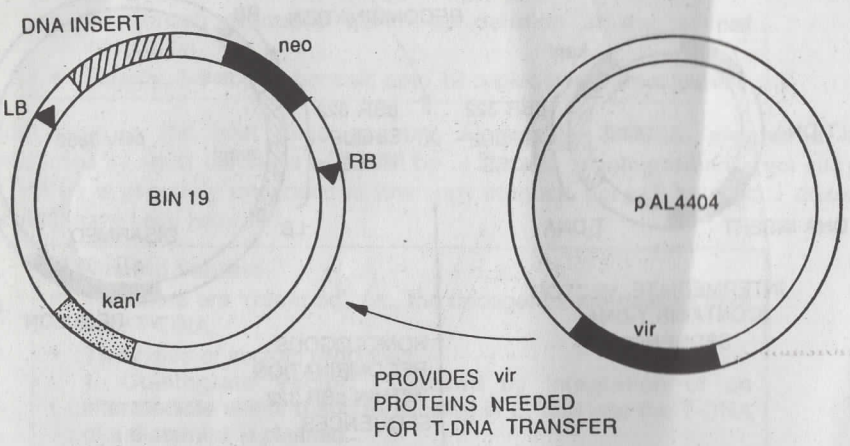


Fig. 37.9. A schematic representation of the binary vectors Bin19 and *p*AL4404 of *p*Ti. Bin19 is derived from the broad host range plasmid *p*RK252; it contains a *neo* marker gene for the selection of transformed plant cells and a polylinker sequence containing the *lacZα* gene within which DNA insertion occurs (to enable the identification of white bacterial colonies containing the recombinant DNA). *p*AL4404 is derived from *p*TiAch5; it contains the *vir* region but its T-DNA has been deleted. LB and RB, left and right borders of T-DNA.

Transformation Methods

1. *Agrobacterium* - mediated.
2. Direct gene transfer:
   - (*i*) Chemical methods, *e.g.*, PEG.
   - (*ii*) Electroporation.
   - (*iii*) Particle gun delivery.
   - (*iv*) Lipofection.
   - (*v*) Microinjection.
   - (*vi*) Macroinjection.
   - (*vii*) Pollen transformation.
   - (*viii*) Delivery *via* growing pollen tubes.
   - (*ix*) Laser-mediated.
   - (*x*) Fiber-mediated.

- Methods 1 and 2 (*ii*) and 2 (*iii*) are the most commonly used.

37.5. TRANSFORMATION TECHNIQUES

A variety of techniques have been used to introduce transgenes into plant cells; these can be grouped into the following two broad categories: (1) *Agrobacterium*-mediated and (2) direct gene transfers.

37.5.1. *Agrobacterium* - Mediated Gene Transfer

Gene transfer through *Agrobacterium* is achieved, usually, as follows. The appropriate gene construct is inserted within the T-region of a disarmed Ti plasmid; either a cointegrate or a binary vector is used. The recombinant DNA is placed in

Agrobacterium, which is then co-cultured for about 2 days with the plant cells or tissues, *e.g.*, leaf discs, to be transformed. In general, the transgene construct includes a selectable reporter gene, *e.g.*, the bacterial *neo* gene. The *neo* gene is linked with suitable regulatory sequences that are functional in plant cells.

During the co-culture, acetosyningone released by plant cells induces the *vir* genes, which brings about the transfer of recombinant T-DNA into many of the plant cells. The T-DNA would become integrated into the plant genome, and the transgene would be expressed. As a result, the transformed plant cells would become resistant to kanamycin (due to the expression of *neo* gene). After 2 days, the leaf discs are transferred onto a shoot regeneration medium containing appropriate concentrations of kanamycin and carbenicillin (Fig. 37.10). Kanamycin allows only transformed plant cells to divide and regenerate shoots in about 3-4 weeks, while carbenicillin kills *Agrobacterium* cells. The shoots are separated, rooted on a kanamycin containing medium, and finally transferred into soil.

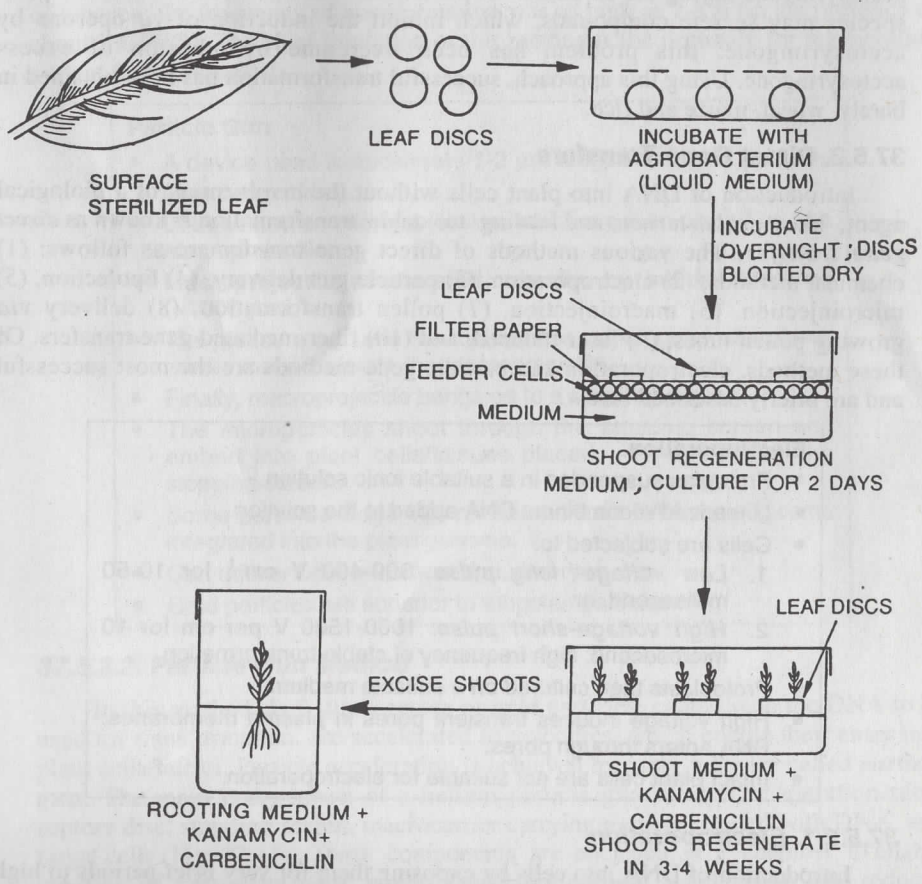


Fig. 37.10. A schematic representation of the generalized procedure for *Agrobacterium*-mediated gene transfer.

Agrobactetrium-Mediated Gene Transfer

- A suitable plant tissue is co-cultured with *A. tumefaciens* (~2d).
- Acetosyringone may be added during co-culture in case of monocots.
- Plant tissues transferred to *selective medium* to eliminate *Agrobacterium* cells and to select transformed plant cells.
- Selective medium also supports shoot regeneration.
- Shoots rooted on a suitably modified selective medium and transferred to soil.
- (*Selective medium* has carbenicillin and a selective agent that selects transformed plant cells.)

Efficient transformation of monocot cells can be obtained by providing acetosyringone during the co-culture of plant cells with *Agrobacterium.* Some plant species may secrete compounds, which inhibit the induction of *vir* operons by acetosyrimgone; this problem has been overcome by addition of excess acetosyringone. Using this approach, successful transformation has been obtained in barely, wheat, maize and rice.

37.5.2. Direct Gene Transfers

Introduction of DNA into plant cells without the involvement of a biological agent, like *Agrobacterium,* and leading to stable transformation is known as ***direct gene transfer***. The various methods of direct gene transfer are as follows: (1) chemical methods, (2) electroporation, (3) particle gun delivery, (4) lipofection, (5) microinjection, (6) macroinjection, (7) pollen transformation, (8) delivery *via* growing pollen tubes, (9) laser-induced, and (10) fiber-mediated gene transfers. Of these methods, electroporation and particle gene methods are the most successful and are briefly described below.

Electroporation

- Protoplas suspended in a suitable ionic solution.
- Linearized recombinant DNA added to the solution.
- Cells are subjected to:
  1. *Low voltage- long pulse*: 300-400 V cm^{-1} for 10-50 millisecond, or
  2. *High voltage-short pulse*: 1000-1500 V per cm for 10 microsecond; high frequency of stable transformation.
- Protoplasts then cultured on a suitable medium.
- High voltage induces transient pores in plasma membranes; DNA enters through pores.
- Intact plant cells are not suitable for electroporation.

37.5.2.1. Electroporation

Introduction of DNA into cells by exposing them for very brief periods to high voltage electrical pulses is called ***electroporation***. The high voltage is believed to induce transient pores in plasma lemma through which DNA enters the cells. There are basically the following two systems of electroporation: (1) low voltage-long pulse method and (2) high voltage-short pulses approach. For tobacco mesophyll

protoplasts, the typical values are as follows: for low voltage-long pulse method, 300-400 V cm^{-1} for 10-50 ms (milliseconds; exponential decay), and for the high voltage-short pulse approach, 1000-1500 V cm^{-1} for 10 μs (microseconds; square wave pulse generators).

Plant protoplasts are suspended in a suitable ionic solution containing linearized recombinant DNA. The electroporation mixture is then exposed to the chosen voltage-pulse combination for the desired number of cycles. Protoplasts are then cultures to obtain cell colonies and plants. Generally, low voltage-long pulse produce high rates of transient transformation, while high voltage-short pulse give high rates of stable transformation. Transformation frequencies are increased several -fold by (1) a heat-shock to protoplasts just prior to electroporation and the presence of low concentrations (about 8%) of PEG during electroporation.

Electroporation has been used to produce stably transformed cell lines and/or plants in several plant species, *e.g.,* tobacco, petunia, maize, rice, wheat, etc. In case of tobacco, the frequency of transformation was as high as 2-8% (in the presence of about 7% PEG). The chief limitation of this method is the necessity for regeneration system from protoplasts.

Particle Gun

- A device used to accelerate 1-2 μm tungsten or gold particles (*microprojectiles*).
- DNA coated on microprojectiles, which are then placed on to a macroprojectile.
- Macroprojectile accelerated, usually, by helium pressure (~1000 p.s.i.).
- The pressure ruptures the rupture disc, which generates shock waves that accelerate the macroprojectile.
- Finally, macroprojectile bangs on to a stopping screen.
- The microparticles shoot through the stopping screen and embed into plant cells/tissues placed, ~10 mm below the stopping screen.
- Some particles lodge into nuclei and their DNA may become integrated into the plant genome.
- Can deliver DNA into shoot tips, embryos, etc.
- Gold particles are superior to tungsten particles.

37.5.2.2. Particle Gun Method

In this method, 1- 2 μm tungsten or gold particles, coated with the DNA to be used for transformation, are accelerated to velocities, which enable their entry into plant cells/nuclei. Particle acceleration is achieved by using a device called ***particle gun***. The main components of a helium particle gun are gas acceleration tube, rupture disc, stopping screen, macrocarrier carrying particles coated with DNA, and target cells (Fig. 37.11). These components are enclosed in a chamber to enable creation of partial vacuum, which facilitates particle acceleration and reduces damage to plant cells. After creation of partial vacuum, sufficiently pressurised helium gas is released in the acceleration tube to break the rapture disc. This generates pressure shock waves, which accelerate the macroprojectile to which the

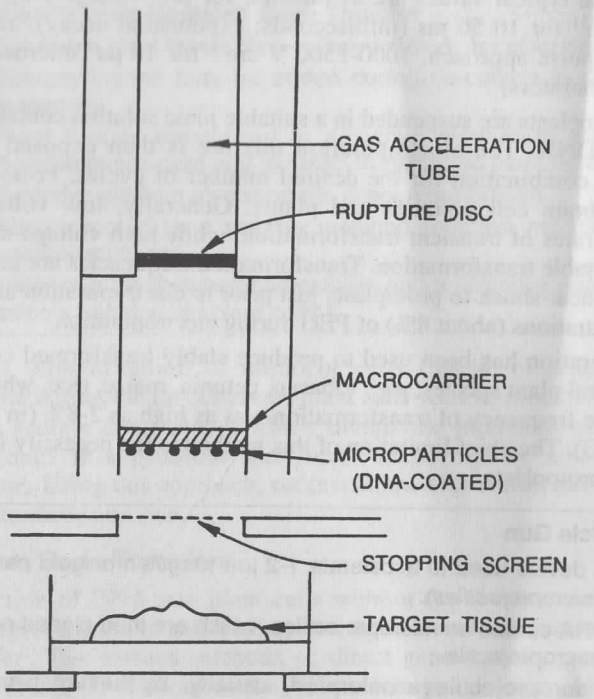


Fig. 37.11. A simplified representation of those components of a helium pressure particle gun device, which are involved in DNA delivery . The microprojectiles are coated with the DNA to be delivered in the target tissue.

DNA-coated microprojectiles are attached. The macroprojectile is stopped by a stopping screen, and the microprojectiles pass through this screen and become embedded into the cells kept about 10 mm below the stopping screen. Helium is preferable to air since it is lighter and offers certain advantages. Generally, 1000 psi (pounds per square inch) of helium pressure is used for acceleration.

This approach, also called ***biolistic technique***, has been used to produce stable gene transfers in cotton, maize rice, sorghum, soybean, sugarcane, tobacco, wheat, papaya, poplar, etc. This gene transfer method is quite attractive because DNA can be delivered into the cells of shoot meristems of shoot tips/embryos, which makes gene transfer independent of regeneration ability of the species. But its chief limitation is the need for a costly specialized acceleration device (particle gun) and the relatively costly consumables.

37.6. INTEGRATION OF TRANSGENES

The main features of transgene integration from direct DNA delivery into the plant genomes are listed below.

1. Integration occurs at random sites in the genome.
2. Linearized DNA is much more efficient in producing stable integrations than is circular DNA.

Transgene Integration

- Integration at random sites.
- Linearized DNA more effective.
- Co-transformation by unlinked DNA molecules when they are used together.
- Multiple copy integrations usually at one site.
- Higher the DNA concentration:
  1. higher the transformation frequency, and
  2. higher the frequency of multiple copy integration.
- Transgenes show Mendelian inheritance.
- Transgenes ordinarily show dominance.
- A heterozygote for a transgene is, in fact, a *hemizygote.*

3. Different unlinked DNAs introduced together usually become linked in the cell and show co-transformation.
4. Multicopy integrations usually occur in tandem at one site.
5. Higher DNA concentrations increase the transformation frequency, but the number of copies integrated per cell also becomes high (5-10); at lower DNA concentrations, the copy number ranges between 1 and 5.

The chief features of gene integration by *Agrobacterium*-mediated gene transfers are as follows: (1) integration occurs in the nuclear genome, only exceptionally in the chloroplast genome, (2) integration occurs at random, (3) one to several integrations may occur in the same nucleus, and (4) multicopy insertions usually occur in tandem arrays.

37.7. INHERITANCE OF TRANSGENES

Stably integrated transgenes are inherited in Medelian fashion, and usually show dominance. But in subsequent generations, some instability may occur probably due to DNA methylation or rearrangements of the T-DNA region. In addition, homologous recombination between copies of the transgene inserted in the same nucleus can also lead to instability of the transgene.

37.8. ANALYSIS AND CONFIRMATION OF TRANSGENE INTEGRATION

The confirmation of transgene integration and analysis of its function is based on several approaches, which have been considered in some detail in Chapter 36 (Section 36.9). In brief, the presence of transgene may be detected by the specific phenotype produced or by assaying the enzyme/protein produced by using various assays, including western blot and ELISA. The transgene integration is confirmed by Southern blot after an individual tests positive for dot blot or PCR.

37.9. EXAMPLES OF USEFUL GENE TRANSFERS

The gene transfer technology has enabled mobilization of genes from any organism or even synthetic DNA sequences into the genome of any plant species (potentially). This capability can be and has been exploited to tailor plant genomes to suit various human needs. Some of the successful examples of gene transfers of economic value are concerned with the following: (1) insect resistance, (2) virus resistance, (3) seed

protein quality, (4) fatty acid composition, (5) disease resistance, (6) biochemical production, (7) suppression of or overproduction by endogenous genes, (8) vaccines, etc. Some of the transgenic crops have reached the market, while others are on their way.

Insect Resistance

- *cry* gene of *B. thuringiensis* has been the most successful.
- The crystal protein is cleaved by insect proteases in midgut; this releases the toxin fragment.
- Toxin binds to specific receptor present on epithelial cells.
- This induces pore formation and cell lysis.
- Truncated *cry* genes expressed in plants generate insect resistance.
- Base sequence modification enhances *cry* gene expression.
- Insect resistant maize, cotton, etc. varieties being grown in U.S.A. and several other countries.
- *cry*I acts against Lepidoptera, *cry*IIA against Lepidoptera and Diptera, *cry*IIB against Diptera, *cry*III against Coleoptera, and *cry*IV against Diptera.
- *Cry* proteins act against larvae of the insects.

37.9.1. Insect Resistance

The efforts to produce insect resistant transgenic plants have been quite successful; insect resistant transgenic varieties (containing *cry* gene) of several crops are already in commercial cultivation in U.S.A. and many other countries. Insect resistance can be produced by either the *cry* gene from the bacterium *Bacillus thuringiensis* or some other protease/amylase inhibitor gene, *e.g.*, the cowpea trypsin inhibitor gene. *B. thuringiensis* has been used since World War I, particularly in Europe, to control some insect pests. The *cry* gene of *B. thuringiensis* produces a protein, which forms crystalline inclusions in the bacterial sopres. These crystal proteins are responsible for the insecticidal activities of the bacterial strains. The *cry* genes have so far been grouped into 4 main groups, *viz.* *cry*I, *cry*II, *cry*III and *cry*IV; these are further subdivided into 16 distinct groups, which either code for a 130 kDa or a 70 kDa protein. These proteins are solubilized in the alkaline environment of insect midgut and are then proteolytically processed to yield a 60 kDa toxic core fragment.

The CryI proteins are insecticidal to Lepidopteran insects. The CryIIA proteins are active against both Lepidoptera and Diptera, while CryIIB is specific to Diptera. The CryIII proteins are active against Coleoptera species, while CryIV proteins are specific to Diptera. The toxic fragments bind to specific high- affinity receptors present in the membranes of midgut epithelial cells (Fig. 37.12). As a result, the membranes develop pores that permit influx into the epithelial cells of ions and water, which causes their swelling and eventual lysis. The specificity of crystal proteins to their target insect species is mainly due to the presence of specific receptors in the midgut epithelium of these insects.

The *cryIA* genes have been successfully transferred into tobacco, potato and tomato. Typically, truncated *cry* genes are used for the production of transgenic

plants since the level of expression of the complete genes in the transgenic plants is extremely low. In addition, the base sequences of the *cry* genes have been modified in such a way that the amino acid sequences encoded by them remain unchanged. These modifications have further increased the expression level of *cry* genes. As a result, a suitably modified truncated version of a *cry* gene may have upto 100-fold increased expression as compared to that of the native *cry* gene. Field tests have been quite successful with transgenic tobacco, maize, cotton, etc.

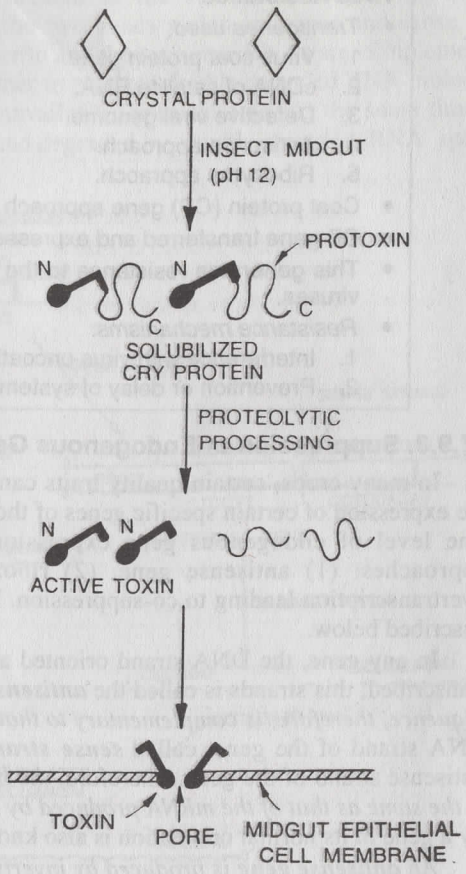


Fig. 37.12. A schematic representation of the production of toxin fragments from Cry proteins by proteolytic processing, and the formation of pores in the midgut epithelial cell membranes. C and N indicate the C-terminal and N-terminal ends, respectively, of the protoxin molecules.

37.9.2. Virus Resistance

Several approaches have been used to engineer plants for virus resistance; these approaches are as follows: (1) coat protein gene, (2) cDNA of satellite RNA, (3) defective viral genome, (4) antisense RNA approach, and (5) ribozyme-mediated protection. Of these approaches, the coat protein-mediated resistance has been the most successful; a virus resistant scquash variety developed by using this approach is in commercial cultivation in U.S.A.

Transgenic plants having virus coat protein gene linked to a strong promoter have been produced in many crop plants, *e.g.*, tobacco, tomato, alfalfa, sugarbeet, potato, etc. It appears to be a common feature that expression of a virus coat protein gene not only confers resistance to the concerned virus but also gives a measure of resistance to related viruses. The effectiveness of coat protein (CP) gene in conferring virus resistance can be affected by both the amount of coat protein produced in the transgenic plants and by the concentration of virus inoculum. Most likely the resistance generated by CP is due to the blocking of the process of uncoating of virus particles, which is necessary for viral genome replication as well as expression. However, other effects seem to be involved in producing coat protein-mediated virus resistance, *e.g.*, prevention or delay of systemic spread of the viruses.

Virus Resistance

- *Transgenes used*:
  1. Virus coat protein gene.
  2. cDNA of satellite RNA.
  3. Defective viral genome.
  4. Antisense approach.
  5. Ribozyme appraoch.
- Coat protein (CP) gene approach is the most successful.
- CP gene transferred and expressed in plants.
- This generates resistance to the virus concerned and related viruses.
- *Resistance mechanisms*:
  1. Interference with virus uncoating.
  2. Prevention or delay of systemic spread of virus.

37.9.3. Suppression of Endogenous Genes

In many crops, certain quality traits can be improved by reducing/eliminating the expression of certain specific genes of those crops, *i.e.*, their ***endogenous genes***. The level of endogenous gene expression can be reduced by the following approaches: (1) antisense gene, (2) ribozyme, (3) gene disruption, and (4) overtranscription leading to co-suppression. The antisense gene approach is briefly described below.

In any gene, the DNA strand oriented as 3′→ 5′ in relation to its promoter is transcribed; this strands is called the ***antisense strand*** (Fig. 37.13). *The mRNA base sequence, therefore, is complementary to that of the antisense strand.* The remaining DNA strand of the gene, called ***sense strand***, is naturally complementary to the antisense strand of the gene. Therefore, *the base sequence of sense strand of a gene is the same as that of the mRNA produced by it.* Hence, the hnRNA/mRNA produced by a gene in its normal orientation is also known as ***sense RNA***.

*An **antisense gene** is produced by inverting,* i.e., *reversing the orientation of the protein encoding region of a gene in relation to its promoter. As a result, the natural sense strand of the gene becomes oriented in the 3′→ 5′ direction with reference to its promoter, and is now transcribed. (The normal antisense strand is not transcribed since now its orientation is 5′→ 3′.)* The RNA produced by the antisense gene has the same sequence as the antisense strand of the normal gene (except for T in DNA in the place of U in RNA); therefore this RNA is known as ***antisense RNA*** (Fig. 37.13).

Endogenous Gene Suppression

- *Endogenous gene* : Gene naturally present in an organism.
- Gene suppression is done to generate the desired phenotype.
- *Approaches* :
  1. Antisense RNA.
  2. Ribozyme approach.
  3. Gene disruption.
  4. Co- suppression.

When an antisense gene is present in the same nucleus as the normal endogenous gene, transcription of the two genes yields antisense and sense RNA transcripts, respectively. Since the sense and the antisense RNAs are complementary to each other, they pair with each other to produce double-stranded RNA molecules. This event makes (1) the mRNA unavailable for translation. At the same time, (2) the RNA double-strand is attacked and degraded by double-stranded RNA specific

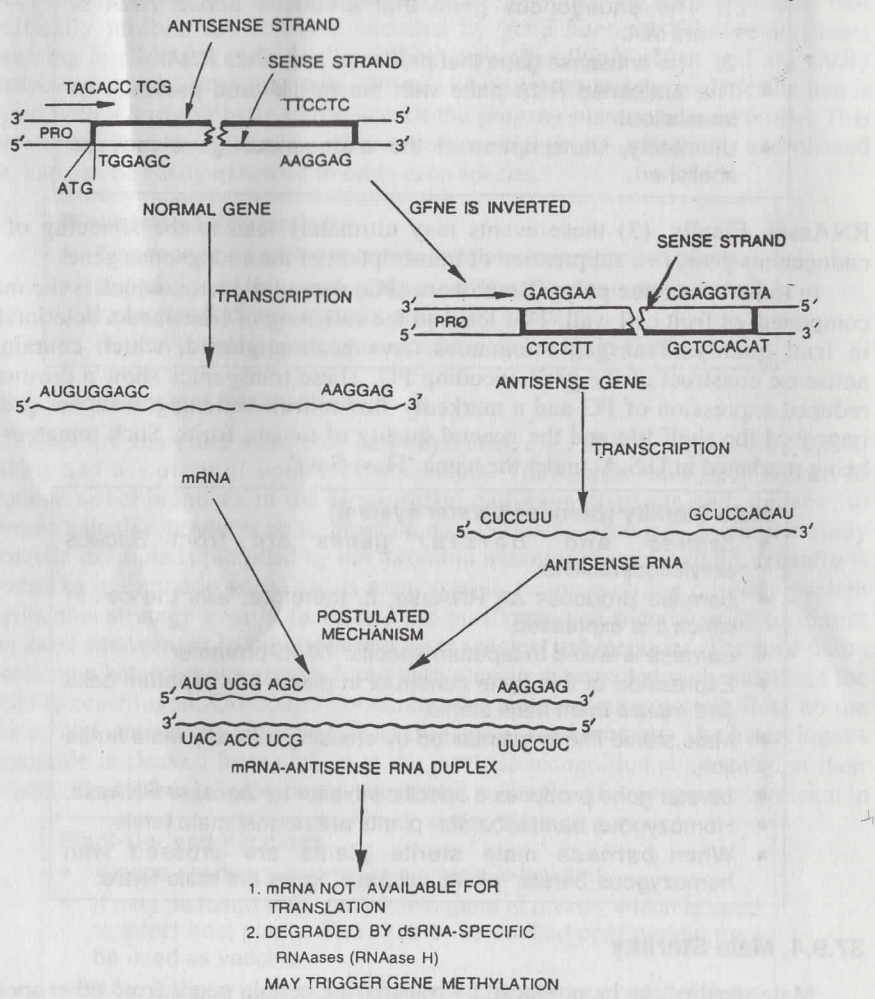


Fig. 37.13. A schematic representation of antisense RNA technology. The antisense gene is produced by inverting the normal endogenous gene (except its regulatory sequences). The mRNA produced by normal gene is complementary to the antisense RNA produced by the antisense gene; as a result, these two RNAs pair to form a duplex, which effectively makes the mRNA unavailable for translation. Antisense strand, the strand of the normal gene that is transcribed, sense strand, the strand of the normal gene that is not transcribed (it has the same base sequences as the mRNA); dsRNA, double-stranded RNA.

Antisense RNA Technology

- *Antisense Gene* : A gene in which the coding sequence is inverted in relation to its promoter; it produces *antisense RNA.*
- Antisense construct of the gene to be suppressed is integrated in the genome.
- *The cell now has*:
  1. The endogenous gene that produces *sense RNA* or *mRNA.*
  2. The antisense gene that produces *antisense RNA.*
- The antisense RNA pairs with the mRNA and prevents its translation.
- Ultimately, transcription of the endogenous gene may be abolished.

RNAases. Finally, (3) these events may ultimately lead to the silencing of the endogenous gene, *i.e.*, suppression of transcription of the endogenous gene.

In tomato, enzyme polygalacturonase (PG) degrades pectin, which is the major component of fruit cell wall. This leads to the softening of fruits and a deterioration in fruit quality. Transgenic tomatoes have been produced, which contain an antisense construct of the gene encoding PG. These transgenics show a drastically reduced expression of PG and a markedly slower fruit softening. This has greatly improved the shelf life and the general quality of tomato fruits. Such tomatoes are being marketed in U.S.A. under the name 'Flavr Savr'.

Male Sterility (*barnase/barstar* system)

- *barnase* and *barstar* genes are from *Bacillus amyloliquefaciens.*
- *barnase* produces an RNAase; it, therefore, kills the cell in which it is expressed.
- *barnase* is linked to tapetum-specific *TA-29* promoter.
- Expression of this gene construct in plants kills tapetum cells and makes them male sterile.
- Male sterile line is maintained by crossing with any male fertile line.
- *barstar* gene produces a specific inhibitor for *barnase* RNAase.
- Homozygous *barstar/barstar* plants are normal male fertile.
- When *barnase* male sterile plants are crossed with homozygous *barstar* plants, all the progeny are male fertile.

37.9.4. Male Sterility

Male sterility can be produced by transferring certain genes from other species, *e.g., rolB* and *rolC* genes from *Agrobacterium rhizogenes, barnase* gene from *Bacillus amyloliquefaciens*, etc. Gene *barnase* encodes an RNAase, which is cytotoxic and kills the cells in which it is expressed. The expression of *barnase* is confined to tapetal cells by fusing it with the promoter of tobacco tapetum-specific gene *TA29* (gene construct: *PTA29-barnase; P* = promoter). When the chimaeric gene construct *PTA29-barnase* was transferred and expressed in tobacco and oilseed rape, the tapetal cells of anthers were destroyed, and there was no pollen development. However, there was no effect on female fertility. Since the male

sterility due to *barnase* is dominant, the male sterile plants are always heterozygous (*barnase/–*; the – sign indicates the absence of *barnase* gene in the homologous chromosome). These male strile plants can be maintained by crossing them to any normal, nontransformed male fertile line (—/—; *barnase* gene absent).

The male fertility of a *barnase/–* male sterile line can be restored by another gene, *barstar,* from *B. amyloliquefaciens.* Gene *barstar* encodes a protein that specifically inhibits the RNAase encoded by gene *barnase.* Transgenic plants expressing *barstar* are male fertile without any phenotypic effect and are easily maintained in the homozygous state. When a homozygous *barstar* male fertile line is crossed with a *barnase* male sterile line all the progeny plants are male fertile. This male sterility/fertility system has shown commercial promise in maize and oilseed rape, and can be easily extended to other crop species.

Biochemical Production

- Transgenes encoding enzymes/ proteins expressed in plants.
- Separation of transgenic protein facilitated by its expression as fusion product with oleosin, *e.g.,* in *B. napus.*
- Hirudin is being produced on commercial scale in Europe.

37.9.5. Biochemical Production

Plants are the chief source of carbohydrates, *e.g.,* starch, sugar, etc., lipids, proteins and a variety of unique biochemicals. Transgenes have been shown to introduce novel branches in the biosynthetic pathways of plants and, thereby, to generate valuable products or to produce new, valuable proteins. A commercially successful example is provided by the thrombin inhibitor protein hirudin. ***Hirudin*** is encoded by a synthetic gene, and is expressed in fusion with the oil-body protein oleosin; this strategy greatly facilitates the purification of heterologous (different from those endogenous to the concerned plant species) polypeptides. The seed tissue expressing a heterologous protein fused with oleosin is extracted with water and the extract is centrifuged. Oil bodies containing the oleosin-fusion protein float on the surface, and are easily separated from the rest of seed proteins. The heterologous polypeptide is cleaved from oleosin at the protease recognition site located at their junction. Thus hirudin provides a successful example of transgene expression in

Plant-Derived Vaccines

- A gene encoding an antigenic protein is isolated.
- It may be fused with coat protein gene of a virus, which is used to infect host plants. The virus or its purified coat protein may be used as vaccine.
- It the antigen produces immunogenic response when delivered orally:
  1. The gene for this antigen is expressed in a suitable plant tissue, *e.g.,* banana fruit.
  2. The tissue is fed to the animal that is to be vaccinated.
  3. This immunizes the animal.
  4. The approach is in advanced stages of development.
- Plant - derived vaccines are expected to be cheaper, and easier to store and handle.

plants for the isolation of a polypeptide of interest on a commercial scale. Production of transgenic proteins in plants offers several advantages. But one of the chief limitations of this approach is the high cost of purification of the transgenic proteins.

37.9.6. Plant-Derived Vaccines

Conventional vaccines consist of attenuated or inactivated pathogens. In case of many pathogens, the gene encoding a critical antigen has been isolated and expressed in bacterial/animal cells, and the recombinant protein so produced is used as a vaccine; such vaccines are called ***recombinant vaccines.*** (A ***recombinant protein*** is encoded by a natural or synthetic gene, which is present in a recombinant DNA construct and is expressed in a host other than the species to which the gene belongs.) But usually the entire protein molecule is not necessary for an effective immunization since the antigenic property is ordinarily localized in a small part of protein molecules. Therefore, either a constituent polypeptide (in case of heteromultimeric proteins) or a small part of a polypeptide may be used as a vaccine; such vaccines are known as ***subunit vaccines.*** Subunit vaccines are produced using recombinant DNA technology for cloning and expressing the appropriate DNA sequences encoding antigenic polypeptides or polypeptide segments in suitable hosts.

Recombinant vaccines are usually produced through bacterial fermentation or in animal cell cultures, which often makes their cost prohibitively high. In addition, storage and transport of vaccines, especially in developing countries, presents many problems, *e.g.*, cold storage. Therefore, plants are being developed as alternative vaccine production and delivery systems. There are two basic objectives of such efforts: (1) development of edible vaccines, and (2) production of recombinant antigenic proteins to be used as vaccines. The strategy for the production of edible vaccines is briefly described below.

Antigens of several pathogens, *e.g.*, enteric pathogens, produce immunogenic response when delivered orally; such antigens are good candidates for edible vaccines. The strategy, in simple terms, is as follows. The gene encoding the orally active antigenic protein is isolated from the pathogen, and a suitable construct for constitutive or tissue-specific expression of the gene is prepared. The gene construct is introduced and stably integrated into the genome of selected plant species, and is expressed to produce the antigen. The appropriate plant parts containing the antigen may be fed raw to animals or humans to bring about immunization. For animals, crops used as feed, *e.g.*, alfalfa, and other forage/fodder crops, are suitable for the expression of such antigens, while for humans fruits like banana, which are consumed raw, have to be used.

Thus ***edible vaccines*** are the produce of transgenic plants in which an orally active antigen of the target pathogen is expressed and accumulated, and which is fed to animals/humans for immunization against the pathogen. The edible vaccines are expected to alleviate the storage problems, offer an easy delivery system by feeding and would have much lower costs than the recombinant vaccines produced by bacterial fermentation. The development of edible vaccines for animals is the first objective, and would be followed by those for humans once the basic mechanisms become well understood. An example of an edible vaccine is provided by the *E. coli* heat labile enterotoxin (LT) B subunit (LT-B) expressed in potato.

37.10. PROBLEMS IN GENE TRANSFER

Gene transfers in plants present several problems; the main problems are summarized below.

1. Many transgenes are expressed at low levels in plants; this usually limits the usefulness of these transgenic plants. Often suitable strategies can be devised to overcome this problem, *e.g.*, in the case of *cry* genes.
2. Frequently, the expression of transgenes declines, or may even be completely suppressed, in the transgenic plants; this is called ***gene silencing***. The machism of gene silencing remains unknown. But, in general, the higher the number of copies per cell of a transgene, the more frequent is its silencing. Stable transgene expression can be obtained by screening a large number of independent transformants. Further, gene silencing can be minimised by a selection for stable transgene expression in the segregating generations of crosses between transgenic lines and elite normal varieties.
3. Frequently, transgenic plants may show some undesirable features like necrosis, reduced growth, sterility, etc. The causes of these problems usually vary, but often suitable methods/ approaches can be devised to resolve these problems.
4. The frequency of transformation is rather low in case of most transformation methods. Some DNA delivery techniques, *e.g.*, electroporation, microinjection, etc., do show relatively high (>1%) transformation frequency, but they require reliable regeneration protocols from protoplasts, high skill and expensive equipment.
5. Gene integration occurs at random sites and often in multiple copies; the latter may often lead to gene silencing. It would be desirable to develop techniques of DNA delivery that give high frequency transformation, single copy transgene integration and possibly targetted gene integration.
6. In case of *Agrobacterium*-mediated gene transfer, selection methods may not always be able to completely eliminate the bacterium. As a consequence, plant tissues may be contaminated by *Agrobacterium* cells carrying the transgene construct. Such plants/tissues will give a false

Biosafety Considerations

- *Fears*: Transgenic plants may:
  1. Become persistent weeds.
  2. Transfer their genes to weeds making them more persistent.
  3. Be detrimental to the environmental.
  4. Be harmful to human health.
- *Testing procedures*: Stepwise evaluation from:
  1. laboratory,
  2. growth chamber,
  3. green-house,
  4. limited field - tests to,
  5. large scale field tests.
- This should allow identification of the risks involved, if any.

positive signal in dot blot assay. But such plants can be detected by: (*i*) assays for transgene expression, (*ii*) Southern hybridization using different restriction enzymes, and (*iii*) using a sequence of Ti/Ri plasmid outside the T- region as a probe for hybridization in the dot blot assay.

37.11. BIOSAFETY REGULATIONS FOR TRANSGENIC PLANTS

The main concerns while taking transgenic plants to the field relate to the possibilities of (*i*) their becoming persistent weeds, (*ii*) of gene transfers from them to other plants making the latter more persistent or invasive, and (*iii*) of their being detrimental to the environment. In general, testing of transgenic plants should progress in a stepwise manner from laboratory to growth chamber, to greenhouse, to limited field testing, to large scale field testing. Many countries have developed their own procedures and policies regulating field tests of such plans. In India, DBT, New Delhi is concerned with the regulation of field testing of transgenic plants.

Further, safety of the produce from transgenic plants is a matter of great concern and considerable debate; a widely acceptable procedure for arriving at decisions is yet to be evolved. The International Food Biotechnology Council has developed a report in an attempt to give a scientific basis to guide regulators in assessing the risk. More than technical feasibility in producing and the usefulness of transgenic plants, it is the consumer acceptance which is going to decide, at least for some time in the future, whether such plants will succeed globally as commercial varieties; the case of slow ripening tomatoes in USA exemplifies this point.

QUESTIONS

1. What are transgenic plants ? Briefly describe their production under the following heads: (*i*) a typical plant gene, (*ii*) promoters/enhancers, (*iii*) reporter genes, and (*iv*) *Agrobacterium* co-culture.
2. Briefly describe the organisation of T-DNA and *vir* region of Ti plasmid. Explain the mechanism of T-DNA transfer into plant genome.
3. What is disarming of Ti plasmid ? Explain the organisation of and gene transfer using cointegrate and binary vectors of *p*Ti.
4. Briefly describe the important methods of transformation of plant cells.
5. Describe the production of insect resistant and virus resistant transgenic plants.
6. List the various approaches for the suppression of activity of endogenous genes. Describe antisense RNA technology in some detail and discuss its application.
7. Plants may be used to produce useful biochemicals and or edible vaccines. Comment with the help of suitable examples.
8. Discuss the various applications of transgenic plants and the problems encountered in their production/utilization.
9. Write short notes on:

(*i*) *cry* gene, (*ii*) antisense RNA technology, (*iii*) *barnase/barstar* system of male sterility, (*iv*) edible vaccines, (*v*) electroporation, (*vi*) particle gun, (*vii*) reporter genes, (*viii*) *vir* region, (*ix*) binary vector, (*x*) cointegrate vectors, (*xi*) Cry proteins, (*xii*) *Agrobacterium*, (*xiii*) T-DNA, (*xiv*) virus resistance, (*xv*) transfer of T-DNA, (*xvi*) transgene integration into plant genomes.

38
Biosafety

38.1. INTRODUCTION

Man has been evaluating plants and microorganisms modified through conventional approaches of breeding, mutagenesis and selection. Further, deliberate introduction of disease-causing microorganisms in carefully planned field experiments is also an accepted practice, which is usually required to obtain valid information about the pathogens and their interaction with the host plants, to identify disease resistant plants and to test the efficacy for control of diseases. But with the development of recombinant DNA techniques, fears have been expressed about the consequences from accidents either in the research laboratory or in the use of recombinant DNA-containing organisms in agriculture and the environment.

Recombinant DNA technology enables man to combine DNA sequences from different sources to create functional DNA molecules with novel properties. These molecules are expressed in various organisms (genetically modified organisms, GMOs), which are then used to achieve biomedical, agricultural and environmental objectives. The recombinant DNA techniques initially raised concerns in the scientific community, and remain a public concern even today. It was feared that genetically engineered microorganisms (GEMs) may disturb the ecosystem and its processes, in which they might be released, in the following two ways. (1) They may rapidly multiply and outcompete the native microbes. (2) They may also transfer genes related to virulence or pathogenesis into the native bacterial populations and, thereby, increase their virulence. Similarly, gentically modified plants could pose biological and ecological risks that may be summed up as follows.

1. Unforseen metabolic changes may result from the genetic modifications. In some cases, this may give rise to toxic or allergic metabolites.
2. Unexpected new susceptibilities to pathogens may arise; this may build up the inoculum of pathogens, which may endanger the ecosystem.
3. The new characteristics may be transmitted to related sexually compatible weed species (in case of plants) and to microbial populations. This may be a problem particularly in cases of herbicide resistance, antibiotic resistance and viral genes.
4. Ecosystems may be disturbed by dispersal, persistence or altered reaction to parasites, symbionts or competitors.

Thus it was perceived that the consequences of gene transfer through recombinant DNA technology are less certain and posed a higher risk than those through conventional methods of breeding. Therefore, the main issue remains as follows: 'Is the method of gene transfer using recombinant DNA technology as safe

as the conventional methods of gene transfer ?' As early as 1987, a committee of the National Academy of Sciences (U.S.A.) concluded that "there is no evidence of unique hazards either in the use of recombinant DNA techniques or in the movement of genes between unrelated organisms, that "risk associated with introduction (of GMOs) poses no risks different from the introduction of unmodified organisms and organisms modified by other methods", and that "assessment of risk should be based on the organism, not the method of engineering." In this chapter, we briefly consider the various aspects of safety required for research on recombinant DNA and the use of GMOs.

38.2. HISTORICAL

Discussions on the possible hazards of cloning recombinant DNA molecules began in the early 1970s. The main concerns focussed on (1) laboratory practices needed to handle serious human and animal pathogens, (2) possibility of creation of 'hybrid organisms' with biological activities of an unpredictable nature, and the escape of 'hybrid organisms' from the laboratory with unpredictable consequences.

These concerns were examined by a committee of National Academy of Sciences (U.S.A.) in 1974. As a consequence, the National Institute of Health, U.S.A. (NIH) established the Recombinant Advisory Committee (RAC) in 1974. In February 1975, a historic international meeting was convened at Asilomar, California; the conclusions from the conference were as follows: (1) certain experiments should be deferred, (*ii*) most of the work on recombinant DNA could proceed with appropriate safety measures, (*iii*) potential risks were assigned to different types of experiments, and (*iv*) such safe bacteria and plasmids that could not survive in the environment if they escaped from the laboratory should be developed.

The first NIH guidelines were prepared in 1975; they were more strict than the recommendations of the Asilomar conference. The guidelines were revised after two years; these had major changes and were much less restrictive. The guidelines specified types of research, often according to the organism, and described the physical and biological principles of containment. In 1977, the NIH prepared an Environmental Impact Statement (EIS), which estimated the possible undesirable impacts of experiments and the likelihood of their occurrence based on probabilities of creating and dispersing potentially hazardous agents.

By 1981, most cloning experiments in *E. coli* K-12, certain strains of *Bacillus subtilis* and *Saccharomyces cerevisiae* were considered exempt from other requirements of NIH guidelines. Eventually, complete exemption was granted for most recombinant DNA research. A major revision of the guidelines was effected in 1982; containment levels were lowered, and experiments that were previously prohibited, including deliberate release of recombinant DNA-containing organisms into the environment, were changed to category requiring review and approval by NIH. In U.S.A., the NIH guidelines are followed by all federal agencies that fund research on recombinant DNA.

38.3. DEFINITIONS

A discussion on biosafety guidelines uses certain terms in some specific sense; some of the important terms are briefly described below. ***Modern biotechnology*** means the application of *in vitro* nucleic acid techniques, including recombinant DNA techniques and direct injection of nucleic acid into cells or organelles, or fusion of cells beyond the taxonomic family that overcomes natural physiological, reproductive or recombination barriers and that are not the techniques of traditional breeding and selection. A ***living modified organisms*** or ***genetically modified organism*** (GMO) means any living organism that contains a novel combination of genetic material obtained through the use of modern biotechnology. Thus ***genetically modified organisms (GMOs)*** are those organisms whose genetic material has been modified by methods that transcend modifications effected by traditional selection and breeding, *e.g.*, organisms that are modified by recombinant DNA methods; the term GMO applies to all organisms, including microorganisms.

Planned introduction describes the process of taking GMOs from contained use to the open environment. ***Contained use*** means any operation, undertaken within a facility, installation or other physical structure, which involves living modified organisms that are controlled by specific measures that effectively limit their contact with and their impact on, the external environment. The terms ***deliberate release, deliberate introduction*** and ***'field testing'*** are often used as synonyms of planned introduction.

Recombinant DNA molecules are defined as either (1) molecules that are constructed outside living cells by joining natural or synthetic DNA segments to DNA molecules that can replicate in a living cell, or (*ii*) molecules that result from replication of the molecules described in (*i*) above. Synthetic DNA sequences that are not expressed *in vivo* as a biologically active polynucleotide or polypeptide are exempt from the NIH guidelines. But genomic DNAs of plants and bacteria that have acquired a transposable element, even if the latter was donated by a recombinant vector no longer present, are not subject to the NIH guidelines unless the transposon itself contains recombinant DNA.

38.4. OBJECTIVES OF SAFETY GUIDELINES

The biosafety guidelines are developed to contribute to ensuring an adequate level of protection in the fields of safe transfer, handling and use of living modified organisms resulting from modern biotechnology that may have adverse effects on the conservation and sustainable use of biological diversity, and to reduce risks to human health. The Cartagena Protocol on Biosafety to the Convention on Biological Diversity has been designed to cover the transboundary movement, transit, handling and use of all living modified organisms that may have adverse effects on the conservation and sustainable use of biological diversity, taking into account risks to human health.

The National Institute of Health (NIH, USA) has developed guidelines for recombinant DNA research with a view to specify the practices for constructing and handling (*i*) recombinant DNA molecules, and (*ii*) organisms and viruses containing recombinant DNA molecules. These guidelines require approval and clearance of any recombinant DNA experiment that requires such approval/clearance from NIH or another Federal agency. All experiments involving the deliberate transfer of recombinant DNA, or DNA or RNA- derived from recombinant DNA into human subjects can not be initiated without submission of the required information to NIH and other specified agencies. NIH funding for recombinant DNA research stipulates that such research must be associated with or sponsored by an institution that

assumes the responsibilities assigned in the NIH Guidelines, and ensures that such research complies with the guidelines.

The ultimate objective of guidelines that regulate research and development activities using recombinant DNA technology should be to minimise risk from such activities and, at the same time, encourage these activities.

38.5. RISK ASSESSMENT

Risk is a function of the level of hazard and the probability of occurrence of the hazard. Thus ***risk*** may be defined as the undesirable consequences of an activity in relation to the likelihood of these consequences being realized. A ***hazard*** is any imaginable adverse effect that can be identified and measured. Once the hazard is identified, one can proceed to determine its seriousness and the probability of its occurrence. The hazards due to GMOs are far more difficult to identify than those, for instance, associated with chemical pesticides.

Risk assessment involves determination of the potential and anticipated adverse effects of the recombinant DNA research to the concerned workers, and of the products of such research on human health and the environment consequent to their accidental or deliberate release (in case of living organisms) or as a result of their consumption. Risk assessment should be carried out in a scientifically sound and transparent manner, and should be in accordance with recognized risk assessment techniques. Risk assessment is not at all easy; it consists of hazard identification, hazard analysis, consequence analysis, risk determination and risk evaluation; and ultimately it is a subjective process.

A lack of scientific knowledge or scientific consensus should not necessarily be interpreted as indicating a particular level of risk, an absence of risk, or an acceptable risk. In cases where there is uncertainty about the level of risk, further information may be sought on specific issues.

38.5.1. Assessment of Risk during Laboratory Research

The assessment of risk during laboratory research is usually done in two steps; (1) initial risk assessment and (2) comprehensive risk assessment. The chief objective of risk assessment at this stage is to decide on appropriate laboratory procedures, and physical and biological containment levels for the proposed research.

38.5.1.1. Initial Risk Assessment

The initial risk assessment is made by the investigator on the basis of the risk group (RG) to which the organism, on which he proposes to conduct experiments, belongs. Organisms are classified into four risk groups based on their potential effects on a healthy human adult. The four risk groups (RGs) developed according to the relative pathogenicity for healthy adult humans are as follows.

1. ***Risk Group 1 (RG1).*** Agents belonging to this group are not associated with disease in healthy humans.
2. ***Risk Group 2 (RG2).*** Organisms included in RG2 are associated with such human diseases, which are rarely serious and for which preventive or therapeutic interventions *are often available.*
3. ***Risk Group 3 (RG3).*** RG3 agents are associated with serious or lethal human diseases for which preventive or therapeutic interventions *may be available.*
4. ***Risk Group 4 (RG4).*** Organisms belonging to RG4 are likely to cause serious or lethal human disease for which preventive or therapentic interventions *are not usually available.*

38.5.1.2. Comprehensive Risk Assessment

Following the initial risk group-based risk assessment, a comprehensive risk assessment should be done. This assessment should take into account the agent itself and the manner in which this agent is to be manipulated. The containment level for the experiment may, in the end, be comparable to that of the RG of the agent, or it may be raised or lowered as a result of the comprehensive risk assessment. This assessment takes into account the following.

1. ***Organism Factors.*** These factors include such features of the organism used for the experiment as virulence, pathogenicity, infectious dose, environmental stability, route of spread, communicability, etc.
2. ***Type of Manipulation.*** Careful consideration should be given to the types of manipulations that are planned for some higher RG agents. For example, RG3 agents like yellow fever viruses, etc. should be handled at a higher containment level for animal inoculation and transmission experiments.

38.5.1.3. Categories of Recombinant DNA Research

NIH guidelines recognise six categories of recombinant DNA experiments, which are briefly outlived below (Table 38.1).

TABLE 38.1

Biosafety levels for introduction of recombinant DNA into organisms of various RG and for experiments with whole animals using organisms belonging to different risk groups (based on NIH guidelines)

Organism of risk group	*Biosafety level*	
	Introduction of recombinant DNA into the organism	*Experiments with whole animals using such organisms**
	Approval by IBC before Initiation is Essential	
RG2	BL2	BL2 or BL2–N
RG3	BL3	BL3 or BL3–N
RG4	BL4	BL4 or BL4–N
	Notice to IBC Simultaneous with Initiation	
RG1	BL1	—

*In these experiments, the animal genome has been altered by stable introduction of recombinant DNA into the germline, or the microorganisms have been modified by recombinant DNA.

1. Experiments that require approval by Institutional Biosafety Committee (IBC), Recombinant DNA Advisory Committee (RAC, a committee of NIH), review, and approval by NIH Director before their initiation.

2. Experiments that require approvals by NIH/OBA (Office of Biotechnology Activities, an organ of NIH) and IBC before they can be initiated.
3. Experiments that require approvals by IBC and Institutional Review Boards, and NIH/OBA registration before they are started.
4. Experiments that require IBC approval before their initiation.
5. Experiments that require notice to IBC simultaneous with their initiation.
6. The last category consists of all such experiments that are exempt from regulation of guidelines. All recombinant DNA molecules that are determined not to present a significant risk to health or the environment are exempt from the guidelines.

38.5.2. Risk Assessment for Planned Introduction

Risk assessment for planned introduction of living organisms modified by recombinant DNA technology is done to identify and evaluate the potential adverse effects of such introductions on the environment, including biological diversity and human health. Risk assessment should be done on a case-by-case basis. Risk assessment takes into account the relevant technical and scientific details regarding the characteristics of the following factors.

1. The biological characteristics of ***the recipient or parental organisms,*** including taxonomic status, centres of origin and of genetic diversity, habitat where they are likely to persist or proliferate, etc.
2. Taxonomic status and the relevant biological characteristics of the ***donor organisms.***
3. Characteristics of the ***vector,*** including its origin, source and host range.
4. Genetic characteristics of the ***DNA insert,*** the function it specifies and/or characteristics of the modification introduced.
5. Identity of the ***GMO,*** and the difference between biological characteristics of the GMO and those of the concerned recipient or parental organisms.
6. Suggested ***detection and identification methods,*** and their specificity, sensitivity and reliability.
7. ***Intended use of GMO,*** including new or changed use in comparison to the recipient/parental organism.
8. Information on the location, geographical, climatic and ecological characteristics, including relevant information on biological diversity, of the ***receiving environment,*** *i.e.,* the environment in which planned introduction is proposed.

It has been estimated that since 19th century, nearly a million planned introductions of microorganisms associated with plants have been done into the environment in the U.S.A. alone; this number would be several million on the global scale. There is no case where the organisms did spread beyond the site of introduction and caused unreasonable adverse economic (*i.e.,* crop loss) or ecological (*i.e.,* species dislocation or displacement) effects in the surrounding ecosystems. It is recognized that accelerated evolution results from plant breeding, which could be detrimental to the environment. Indeed civilization itself is seen ecologically disruptive (*e.g.,* construction of roads, factories, airports, etc.). But we can not imagine a world today either without plant breeding or the activities related to civilization. Similarly, vaccination, which requires introduction of pathogens into

animals and humans, is widely accepted in view of its unquestioned benefits. As a result, public is generally quite favourably disposed to genetically modified organisms used as vaccines. The experience so far indicates that there is little to suggest that the new biotechnology processes/products are any more or less safe than the more traditional ones.

In case of transgenic plants, the following are the main concerns related to their planned introduction.

1. ***The likelihood of a transgenic becoming a weed.***
2. ***Likelihood of the transgene being transferred to wild relatives.*** Transfer of transgenes conferring herbicide resistance would be rather hazardous, especially if the movement occurred to a wild, weedy relative, than in case of transgenes specifying, say, insect or disease resistance.
3. It has been suggested that introduction of novel genes may cause perturbation of the ecosystem. The findings of various studies tend to suggest that it is not genetic engineering *per se*, that determines potential effects; rather it depends on the specific combination of genes, crops and locations.
4. Virus-resistance genes may provide opportunities for evolution of newer virulent strains by recombination between these viral genes and those of other infecting viruses for which these plants are not susceptible.
5. Insect resistant plants, e.*g.,* those containing *cry* gene, would be cultivated on a large scale; this may lead to the evolution of insect biotypes resistant to the toxic effects of the Cry proteins. The fate and long-term effectiveness of *cry* genes remains to be determined.
6. It is also feared that transgenic plants may affect the flora and fauna of their phyllosphere as well as rhizosphere.
7. The product of transgenes may pose health hazard, *e.g.,* allergy, to human consumers.

Molecular biologists and ecologists represent diagnally opposite views not only on the perception of risk but also about the manner in which risk should be assessed. Molecular biologists view modern biotechnology as an extension of classical genetics, and believe that the risk can be evaluated by studying the features of transgenes and host organisms. But ecologists argue that there exist many uncertainties concerning the risks, and they emphasize the study of entire ecosystem in which biotechnological products, especially recombinant DNA products, are to be released. The initial response of ecologists appeared to be that planned introductions should never be made unless everything is known about the organism and its role in the ecosystem. But this position gave way to a cautious planned introduction based on scrutiny of several attributes of organisms and environments for risk evaluation.

38.5.3. Risk Assessment of Biotechnology Products

In U.S.A., foods and additives developed by biotechnology are subject to the same rigorous safety standards as are conventional foods and additives, and there are no special requirements for such products. The foods and additives obtained from genetically modified organisms are called ***genetically modified food*** (***GM food***). Risk assessments begins on the basic premise that food derived from such varieties of all food crops that have been developed by conventional plant breeding are

generally safe. Assessment of risk from food obtained from transgenic varieties of crops takes into consideration the following: (1) the food crop that has been modified, (2) the potential for any introduced DNA to encode harmful substances, (3) the safety of proteins encoded by transgenes, and (4) proof that known plant toxicants and important nutrients are within acceptable levels in the new variety. There is some chance that proteins encoded by transgenes may cause allergies in some individuals; this is particularly so for proteins derived from such foods as milk, eggs, wheat, fish, tree nuts and legumes. There are some instances where a transgenic variety had to be withdrawn due to severe allergic reactions in some people, *e.g.*, a soybean transgenic variety in Mexico.

Regardless of the process used to develop a product, the product must be safe for consumption. Information on the following four aspects of the production process and the product would help in evaluation of the safety of a product generated by recombinant DNA technology.

1. Genetic backgrounds of the host and the donor organisms.
2. The technical procedures used for developing the products.
3. Composition of the food, including any potential toxicants and significant nutrients.
4. Relevant toxicological data on the host organism.

Japan has announced mandatory evaluation of GM food for potential health risk from April 2001; it has also been decided to 'label' those GM foods that are found to be 'safe'. Similarly, European Union has adopted legislation requiring mandatory labelling of all GM food. In contrast, U.S. Food and Drug Administration (FDA) requires the food to be labelled as genetically modified only if the changes introduced through genetic engineering have an impact on the safety or nutrition of the food itself. Labelling is considered desirable because it gives the consumer the choice of deciding if he wishes to have GM food, but companies fear that labelling may frighten the consumers. But some companies like Monsanto favour labelling, and in some cases, *e.g.*, in case of a tomato puree made from transgenic tomato, labelling actually promoted the sale initially.

38.6. RISK REGULATION

Subsequent to risk assessment, practices/procedures, assays/tests, biological agents, equipments, and working designs and structures are developed to minimise the risk; this is termed as ***risk regulation.*** Risk regulation can be either (1) reactive or (2) proactive. ***Reactive risk regulation*** develops gradually in a piecemeal manner, and is designed to prevent the proven adverse effects shown by earlier products/activities. Decisions about the need for regulation and the level of regulation are taken on the basis of relevant benefits and costs. The assays for predictive toxicology, acute and chronic toxicity, mutagenicity and teratogenicity are a part of such a regulation; they are designed due to realized hazards in the past and not on the basis for expected risks.

But ***proactive risk regulation*** is developed to avoid potential hazards that are predicted in advance before there is any empirical evidence for such hazards. Assays and trials are, then, designed to evaluate the safety of the product/process in question and to demonstrate absence of the predicted hazards. Therefore, safety issues and the formulation of regulatory system become an integral part of the development of the

industry/product itself. It has been argued, with considerable justification, that a proactive approach is far more desirable than a reactive approach of risk regulation.

38.7. CONTAINMENT

Biological safety in laboratories is achieved by a combination of mechanisms that are divided into the following three categories.

1. A strict adherence to a set of standard laboratory practices that are generally used in microbial laboratories is the basic principle of containment. These practices include aseptic techniques, and a good knowledge of biology of the experimental organisms; all personnel involved in recombinant DNA research must be adequately trained in them.
2. Special procedures, equipment and laboratory installations that provide physical barriers to accidental release of organisms and to prevent contamination or infection of laboratory workers and of people outside the laboratory; these constitute ***physical containment.***
3. In case of experiments on recombinant DNA, specific vectors and hosts are used to reduce their dissemination and survival in the environment; this is known as ***biological containment.***

Thus ***containment*** may be defined as a combination of laboratory procedures, laboratory equipments and installations, and host-vector systems designed to minimise accidental release of organisms during laboratory operations, their dissemination and survival in the environment, and accidental contamination or infection of laboratory workers and of persons outside the laboratory.

38.7.1. Physical Containment

Physical containment consists of the use of special laboratory design, containment equipments and special operational procedures to restrict the number of organisms accidentally released during normal laboratory operations and to prevent contamination or infection of laboratory workers. Physical containment is grouped into four categories; initially, they were designated as P1, P2, P3 and P4, but they are now referred to as BL1 (Biosafety Level 1), BL2, BL3 and BL4, respectively. BL1 is applicable to nonpathogenic organisms, while BL4 represents the most stringent containment level. As the biosafety level increases, the laboratory design and facilities, and the operational procedures become more and more rigid and stringent so as to minimise the risk of accidental release of the GMOs as well as infection of laboratory workers. The guidelines provide a detailed description of the laboratory design, equipments, etc. for the various biosafety levels.

An emergency plan describing the procedures to be followed in case of accidental contamination of personnel or the environment must be prepared. It is the responsibility of the principal investigator that everyone in the laboratory is familiar with both the potential hazards of the work and the emergency plan. In case the work involves a known pathogen, appropriate vaccine, if available, should be made available to all the workers. Further, serological monitoring, when clearly appropriate, must be provided.

Emphasis is placed on ***primary physical containment,*** which is provided by laboratory practices and containment equipment. Special laboratory design provides

secondary physical containment for protection against accidental release of organisms into the environment; it is used where experiments of moderate to high potential hazard are performed. Several different combinations of laboratory practices, containment equipment and special laboratory design may be appropriate for containment of specific research activities.

38.7.2. Biological Containment

Biological containments specifically aims at making such genetic changes in GMOs that reduce the hazard from these organisms when they are accidentally released from laboratories or accidentally transported by workers. Initially, NIH guidelines required the use of *E. coli* strains and vectors that were severely debilitated, *viz.*, that could not infect humans, did not survive and spread outside the laboratory, and could not transfer the introduced foreign genes readily into other organisms in the environment. These objectives can be achieved by using the following: (1) auxotrophic mutants of *E. coli* (limits survival following escape of bacteria), (2) *recA*$^-$ strains (they lack recombination function), (3) plasmid vectors that are non-selftransmissible and non-mobilizable (such vectors can not be transmitted to other bacteria), and (4) *E. coli* strains not carrying a transposon or antibiotic resistance gene. Thus biological containment is based on the vector (plasmid, organelle or virus) used for the construction of recombinant DNA, and the host (bacterial, plant, or animal cell) in which the vector is propagated in the laboratory; therefore it is also called ***HV system*** (***host-vector system***). Any combination of vector and host, which is to provide biological containment shall be chosen or constructed so that the following types of 'scapes' are minimized : (*i*) survival of the vector in its host outside the laboratory, and (*ii*) transmission of the vector from the propagation host to other nonlaboratory hosts.

38.7.2.1. Biological Containment for Prokaryotes

The host-vector systems for prokaryotes are grouped into the following two categories : (1) host-vector 1 systems (HV1) and (2) host-vector 2 systems (HV2).

1. ***Host-vector Systems 1 (HV1).*** In case of *E. coli*, the host is always strain K-12 or a derivative of this strain; hence it is generally termed as ***E. coli K-12 host-vector 1 (EK1) system.*** The vectors under this system include nonconjugative plasmids, *e.g.*, *p*SC101, Col E1 or derivatives thereof, and variants of bacteriophages, such as, phage λ. The *E. coli* K-12 hosts shall not contain conjugation-proficient plasmids, whether autonomous or integrated, or generalized transducing phages.

Other host-vector 1 systems have to be, at a minimum, comparable in containment to *E. coli* K-12 with a nonconjugative plasmid or bacteriophage vector, (Table 38.2). A host-vector system other than *E. coli* K-12 host-vector system can not be designated as HV1 system or HV2 system unless it has been certified by Director, NIH. This certification is based on the construction, properties and test of the HV system by a committee. Certification of a host-vector system does not cover the modifications of either the host or the vector component of that system.

2. ***Host-vector System 2 (HV2).*** HV2 systems provide a high level of biological containment as demonstrated by data from suitable tests performed in the laboratory. Escape of the recombinant DNA either *via* survival of the organisms or *via* transmission of the recombinant DNA to other organisms should be less than 10^{-8}

under specific conditions designed to represent the natural environment. Examples of HV2 host-vector systems are listed in Table 38.2.

38.7.2.2. Biological Containment for Plants

In case of plants, biological containment procedures aim at minimising the risk of dispersal of pollen and seeds of the plants, and of dissemination and establishment of plant-associated microorganisms and insects; these are briefly considered below.

TABLE 38.3

Examples of certified HV1 and HV2 host-vector systems (NIH Guidelines)

Organism	*Host-vector 1 systems*		*Host-vector 2 systems*	
	Host	*Vector*	*Host*	*Vector*
E. coli	Strain K-12 or derivative	*p*SC101, Col E1, or its derivatives	K–12 strain chi-1776	Plasmids *p*BR332, *p*BR325, *p*SC101, pBR327, *p*HB1, etc. Phage vector Charon 3ADP50 or DP50*supF*, Charon 21A, DP50*supF*, etc.
Bacillus subtilis	Strain RUB331 and BGSC1S53	Plasmids *p*UB110, *p*C194, *p*S194, *p*AB124, etc.	ASB29	Plasmids *p*UB110, *p*C194, *p* S194, *p*AB124, etc.
Saccharomyces cerevisiae	—	—	Strains SHY1, SHY2, SHY3 and SHY4	Plasmids YE*p* 2, YE*p*4, YE*p*5, YE*p*6, YE*p*7, etc.
Neurospora crassa	*Strains 37102, 37401, 46316, 64001, etc.	—	—	—
Streptomyces	Species *S. coclicolor, S. lividans, S. parvulus, S. griseus*	Plasmids SCP2, SLP1.2, *p*IJ101, actinophage phi C31 and their derivatives	—	—
Pseudomonas putida	Strain KT2440	Plasmid *pKT*262, *pKT* 263 and *pKT*264	—	—

1. ***Biological Containment Practices for Plants.*** Dispersal of pollen and/or seed produced by plants can be effectively prevented by using male sterile strains.

2. ***Biological Containment Practices for Plant-Associated Microorganisms.*** Effective dissemination of microorganisms beyond the greenhouse can be prevented by using microorganisms that have been genetically disabled to survive in nature and whose natural mode of transmission requires injury to the host.

3. ***Biological Containment Practices for Plant Associated Arthropods.*** Effective dissemination of arthropods and other small animals can be prevented by (1) using nonflying, flight-impaired, nonmotile or sterile arthropods/small animals, and (2) using such animals that have an obligate association with a plant that is not present within the dispersal range of the organism.

The above objectives can also be achieved by some cultural practices and physical containment procedures.

38.8. PLANNED INTRODUCTION OF GENETICALLY MODIFIED ORGANISMS (GMOs)

Planned release of GMOs in the environment is regulated in a proactive manner. But there have been forceful arguments that such regulation should be product-based rather than process-based. In a ***process-based approach***, all GMOs, are considered to pose potential hazards, some of which may be unforeseable, to the environment; they should be regulated in a proactive manner so that any products that may give rise to hazards are excluded from further commercial development. But in ***product-based approach***, all GMOs are not recognized as a special category because they present no environmental hazards beyond those already identified for existing products; therefore, GMOs can be regulated by the reactive systems already in existence. Biotechnology companies favour a product-based regulation for obvious reasons. But the proactive approach to pollution control can greatly stimulate environment-friendly innovations; needless restrictions however must be avoided.

38.8.1. Fields Trials with Genetically-Modified Plants

In case of plants, conventional breeding, cell and tissue culture and genetic engineering share some major similarities in the method of generation, the magnitude and type, and the methods of detection of the genetic changes, but they do differ in some important features. Usually, products of both tissue culture and genetic engineering are usually subjected to few to many back-crosses and selection, and before that they both are subjected to tissue culture. The risk of genetic engineering products lies in the possible unexpected effects of the exotic genes in the new cell environment and the new ecological surroundings.

The type of regulation necessary during field trials should depend on the ability of modified plant to survive, disperse, reproduce, and hybridize with crops and wild plants, and on our ability to confine the plants to the test site. In case of cross-pollinated species, isolation by distance is rarely sufficient to prevent inter-population mating; it has been suggested to introduce male-sterility to prevent interpollination. The available evidence indicates that vectors and regulatory sequences used for the production of transgenic plants do not pose the risk of making the plant a pest. Similarly, a fraction of a coding sequence from a plant pathogen does not confer pathogen status to the plant.

In case of viral coat protein gene-mediated virus resistance, fears of another virus (that infects the transgenic plant) acquiring the engineered protein coat and thereby produce hybrid viruses seems unrealistic. This has been shown by field trials with potato carrying the gene for potato leaf roll virus (PLRV) coat protein; potato viruses S, X and Y failed to acquire PLRV protein coat.

Transgenic varieties of crops expressing *cry* genes (*cry*stal protein gene of *B. thuringiensis*) for insect resistance are in commercial cultivation. Cry proteins are rapidly degraded by the stomach juices of vertebrates, but they could have harmful effects on nontarget insect species, *e.g.*, honeybees (since they collect and feed on the pollen). Oilseed rape (*B. napus*) expressing CpTI (cowpea trypsin inhibitor), chitinase and β-1, 3- glucanase has been used for assaying the impact of these genes

on honey bees. Chitinase did not affect learning performance, while CpTI and β-1, 3-glucanase had detrimental effects.

In general, the effects of transgenic crop varieties on the incidence and dynamics of natural enemies of insect pests appear to be negligible/comparable to those of nontransgenic varieties of the same crop, *e.g.*, in case of transgenic tobacco and cotton [expressing *cry* IA(c)], potatoes, etc. But in some cases, *cry* gene expressing transgenic varieties may lead to adverse effects on predators, *e.g.*, on chrysopid larvae predating on caterpillars fed on *cry* gene expressing maize (reduced fitness), lady bird beetles predating peach-potato aphids fed on potatoes expressing a lectin gene (reduced reproduction rate), etc. A careful evaluation of toxin genes for insect resistance on nontarget insects is warranted since it seems to be a function of the toxin gene, crop species, insect pest and the nontarget insect species.

It has been reported that Cry proteins leak into the soil from transgenic maize roots, where they persist for weeks (much over 3 weeks). The Cry proteins become associated with soil clay/humus, which makes then nonavailable for microbial degradation. It has been suggested that the Cry protein could protect plants from insect pests that damage roots of crops, but it could also harm benign soil organisms or select for insect pests' resistance to itself.

38.8.2. Planned Introduction of Genetically-Engineered Microorganisms (GEMs)

There is some data available from field-research using GEMs. The GEMs that are used live in the environment include biopesticides, nitrogen fixers, vaccines and bioremediation agents that degrade chemical waste. Immunofluorescence techniques have revealed a much higher survival of microbes than was previously suggested by the conventional techniques. Therefore, use of biological containment systems based on lethal genes, possibly coupled to the regulatory system of the substrate to be degraded has been advocated.

GEMs may be constructed that contain active biological containment systems that minimize the survival and dissemination of the GEMs themselves or that prevent the dissemination of recombinant DNA outside the target area. An example of such a biologically contained GEM is provided by *Pseudomonas putida* strain engineered to degrade alkyl-benzoates. This strain contains the *E. coli* gene *gef*; Gef protein causes a collapse of the membrane potential, which leads to cell death. The *gef*⁻ containment system consists of the following elements. (Fig. 38.1).

1. The *xylS* regulator gene, which is expressed constitutively.
2. The regulator gene *lacI* linked to the promoter *Pm*.
3. The structural gene *gef* driven by *E. coli* promoter *Plac*.

These elements function as follows to establish the biological containment system (Fig. 38.1).

1. Gene *xylS* is expressed constitutively; it encodes an inactive regulator protein that becomes active on interaction with *m*-methylbenzoate (3MB).
2. In the presence of 3MB, therefore, the *xylS* regulator is active, and it activates transcription of gene *lacI* driven by *Pm* promoter.
3. The *lacI* repressor, in turn, binds promoter *Plac* and prevents transcription of the *gef* gene linked to it.

Therefore, in the presence of 3MB, transcription of the *gef* gene is repressed and, as a result, Gef protein is not produced. Consequently, *P. putida* cells containing this *gef* system survive normally as long as 3MB is available to them.

4. However, when 3MB is not available to a cell, the *xylS* activator protein remains inactive.
5. As a result, regulator gene *lacI* (linked to the *Pm* promoter) is not transcribed and *lacI* repressor is not produced.
6. In the absence of *lacI* repressor, *gef* gene (driven by *Plac* promoter) is expressed, and Gef protein is produced. Gef causes collapse of the membrane potential, which leads to cell death.

Therefore, the cells of this genetically-engineered *P. putida* strain commit suicide as soon as 3MB is exhaused and becomes unavailable to them; this minimizes their survival in the environment after they have degraded the alkylbenzoates present in the environment.

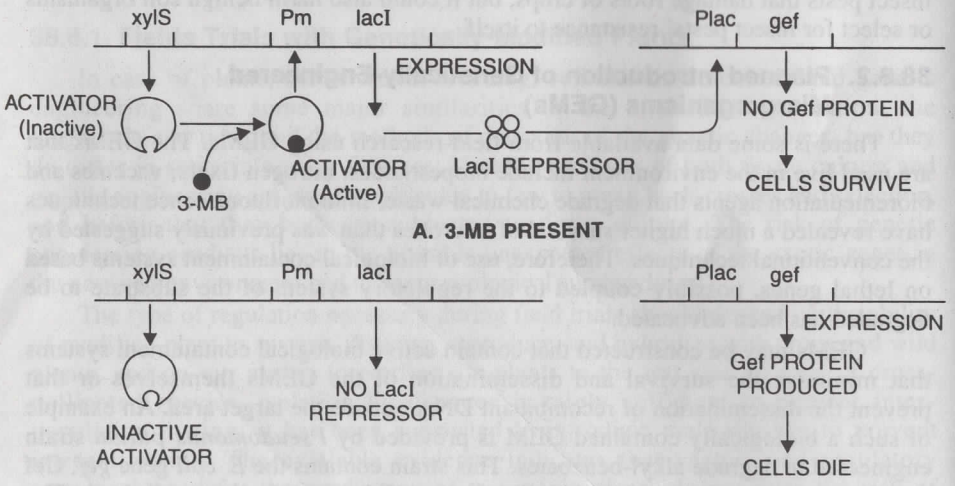


Fig. 38.1. The *gef*-based biological containment system used to construct alkylbenzoate degrading strains of *P. putida*, *e.g.*, strain EEZ30; the biodegradative capability is genetically engineered. 3-MB, *m*-methylbenzoate; *xylS*, produces inactive activator of promoter *Pm*; XylS protein becomes active on interaction with 3-MB; *lacI*, encodes LacI repressor; *Plac*, *E. coli lac* promoter repressed by LacI protein; *gef*, encodes Gef protein, which causes collapse of membrane potential leading to cell death. The *xylS* and *Pm* sequences are from the TOL plasmid (for *meta*-pathway).

The first release of a recombinant microorganism was done in 1986; it related to a Tn5-containing strain of *Pseudomonas fluorescens* in the Netherlands. Since then more than two dozen planned introductions have been done. Most recombinant microorganisms consisted of only a simple inactivation of genes by transposon insertion, genetic tagging of bacteria with specific markers or deletion of genes by molecular genetic procedures *in vitro*. The initial results from these studies suggest

that planned introduction of recombinant bacteria can be done in a safe and responsible manner.

Thus survival of GEMs in nature and the gene transfer from GEMs to native bacterial populations can be minimized by the following strategies.

1. ***Use of RecA⁻ strains;*** RecA protein functions in homologous recombination and repair of DNA damage. Therefore, RecA⁻ strains lack in recombination and show reduced survival.
2. ***Use of auxotrophic mutants*** since they cannot survive in nature.
3. ***Integration of the transgene into the bacterial chromosome*** because plasmids may be transferred to other bacteria.
4. ***Use of appropriate lethal genes, e.g., gef,*** to limit the survival of GEMs to the specified niches.
5. ***Use of appropriately regulated endonuclease gene*** to limit GEM survival and also to prevent transformation by GEM DNA after cell death.
6. ***Use of stable transposon vectors*** (rather than plasmid vectors) that are unable to self-replicate and lack the transposase gene (this limits their transposition to other bacteria).

38.9. BIOSAFETY DURING INDUSTRIAL PRODUCTION

Many of the unit processes of manufacture have the potential to generate biohazards (Table 38.3). Therefore, it is not surprising that several noninfectious illnesses are associated with biotechnology industrial products/processes (Table 38.4). Inhalation appears to be the most significant mode of entry of microbes or microbial products into the body, while entry by ingestion or skin contact can readily be prevented by good standards of operating practice. Most reported health problems are associated with downstream processes, and the risk of allergic reaction is the greatest after the product is more concentrated. During extraction of intracellular enzymes, the greatest risk is posed by the processes of handling the large quantities of cell debris.

An effective monitoring strategy should involve environmental assessment of emissions at all stages of a manufacturing process, and a critical assessment of the biosafety performance of process equipments; in addition, it should maintain adequate surveillance of worker health.

TABLE 38.3

The potential biohazards from process of manufacture in biotechnology industries

Operation	*Hazard*
Raw material weighing, mixing, etc.	Allergic dusts or aerosols
Bioreactor fermentation	Aerosols from the reactor, spillage, effluent contamination of gases
Biomass separation by centrifugation, filtration	Aerosols, spillage, leakage
Product purification	Aerosols, spillage
Product blending, filtration, drying, packaging	Aerosols, dust, spillage
Effluent sterilization, disposal	Discharge of untreated effluent

TABLE 38.4
Selected examples of Noninfectious illnesses associated with production processes in biotechnology industries

Product/Process	*Non-infective illness*
Antibiotic	Cardiovascular disorders, candidiasis etc.
Brewing	Dermatitis, malt fever
Citric acid	Asthma, bronchitis
Enzymes	Asthma, conjunctivitis, dermatitis
Endotoxin	Flu symptoms
Fungal fermentation	Asthma, bronchitis
Single cell protein	Allergic, asthma, dermatitis
Steroids	Feminization of males, hypertension, increased body weight

38.10. BIOSAFETY GUIDELINES IN INDIA

In India, the Ministry of Environment and Forests (MOEF) promulgated in December, 1989 the rules and procedures for the manufacture, import, use, research and release of GMOs as well as products made from such organisms; this was done under the provisions of the Environment Protection Act, 1986 (EPA). Any violation and noncompliance, including nonreporting of activities, in this area would attract the punitive actions provided under the EPA. In addition, The Indian Recombinant DNA Safety Guidelines and Regulations have been prepared by (initially in 1990; revised in 1994) and are available on request from Recombinant DNA Advisory Committee (RDAC), Department of Biotechnology (DBT), New Delhi. Field trials using transgenic plants began in 1995, while China became the first country to begin commercial use of virus resistant transgenic tobacco and tomato in the early 1990s.

The DBT implements the research and development experiments utilizing GMOs and recombinant DNA products, while MOEF implements the large scale commercial use of these. The salient features of these guidelines and regulations are as follows:

1. Every organization involved in research and development using recombinant DNA technology is required to set up an Institutional Biosafety Committee (IBC), which has a DBT nominee. IBC is the nodal point for interaction of the organization with the Government. It is the competent authority to which experiments that are likely to have biohazard potential, and genetic manipulation of plants under containment have to be reported. It also authorizes interstate exchange of etiologic agents, diagnostic specimens, etc. IBC will provide half-yearly reports on the ongoing projects to RCGM.
2. The DBT has a Review Committee for Genetic Manipulation (RCGM), which reviews all the approvals of ongoing projects on GMOs and several other issues related to recombinant DNA research and development. Based on the recommendations of RCGM, trial permits are issued by DBT.
3. Each state has, in addition, a State Biotechnology Coordination Committee, and a District Level Committee, which are also involved, alongwith RCGM, in the inspection and monitoring of the experiments at the field sites.

4. The MOEF has an interministrial committee called The Genetic Engineering Approval Committee (GEAC), which has subject specialists as members, and is the competent authority to decide on the large-scale use of GMOs. The state and district level committees also participate, along with GEAC, in the watch on the field applications of GMOs. GEAC would also involve the user Ministries and the other regulatory authorities, *e.g.*, Directorate General of Health Services in the case of drugs and pharmaceuticals, Ministry of Agriculture for transgenic plants and other organisms related to agriculture, etc.
5. The guidelines recognize 4 levels of risk in the case of experiments with microorganisms, based on pathogenicity of the microorganism, local prevalence of the concerned disease and of epidemic causing strains in India.
6. Experiments with microorganisms, plants and animals are grouped into the following three categories : (*i*) exempt category (for self-cloning experiments), (*ii*) category requiring intimation of initiation to competent authority (*e.g.*, experiments involving nonpathogenic DNA-vector systems), and (*iii*) category requiring review and approval by the competent authority (*e.g.*, cloning of genes for toxins, antibiotic resistance, etc.).
7. Four different *biosafety levels* are recognised and containment facilities for each level are recommended for necessary safeguards.
8. ***Physical containment*** envisages to limit the spread of dangerous microorganisms by (*i*) good laboratory practices, (*ii*) safety equipment, and (*iii*) laboratory design and facilities.
9. ***Biological containment*** consists of the use of vectors and hosts in such a way so that it (*i*) can limit the infectivity of vector to specific hosts, and (*ii*) control host-vector survival in the environment.
10. Experiments with plants are to be done under two types of special environment that may be achieved by using glass-house containment. ***Glass-house containment A*** is used when no plant pathogens are used, and where nonpathogenic vector systems and regeneration from single cells are used. ***Glass-house containment B*** is recommended for experiments involving (*i*) genetically manipulated plant pathogens, including plant viruses, and (*ii*) the growth of plants regenerated from cells transformed by genetically manipulated pathogen vector systems, which still contain the pathogen. Inspection of a 'glass-house containment B' facility by IBSC will be required before approval.
11. Application for recognition of research facility to carry out genetic manipulation should be made to the Department of Environment before the commencement of work.
12. All products obtained using recombinant DNA technology shall be subject to the general regulations normally applicable for such products.
13. The controlled release of GMOs should be done under appropriate containment facilities to ensure safety and to prevent unwanted release in the environment.

14. Prerelease tests of GMOs in agriculture should include elucidation of requirements for vegetative growth and persistence and stability in small plots and experimental fields for years.
15. Planned field experiments with transgenic plants are permitted only after a step-wise (laboratory to growth chamber and greenhouse) evaluation, either in India or elsewhere, to generate data on the following: (*i*) promoter sequence, (*ii*) target gene sequence, (*iii*) regulatory mechanism utilized in the expression cassette, (*iv*) diagram of the expression cassette to describe fully the marker genes used, (*v*) cell lines used for shuttling and amplification of the cassette and several other details of the procedure, and the consequences thereof, used for production of the transgenic plants. In addition, the following is also insisted upon.
    (*a*) Laboratory data to show that the protein products of transgenes are safe to the environment and human beings.
    (*b*) Isolation distance applicable to foundation seed of the crop be provided all round the field having transgenic crops.
    (*c*) A few rows of a nontransgenic line of the same crop as the transgenic one should be planted beyond the isolation distance to act as pollen trap.
    (*d*) Nontransgenic plants should be grown within the isolation distance at 1 or 5 m intervals to determine the distance of pollen escape.
    (*e*) All the vegetative plants and left-over seeds must be destroyed by burning after the experiments are concluded.
    (*f*) After the experiment, the land may be left fallow and all the plants that emerge must be destroyed.
    (*g*) The experimental field may be visited by the company authorized personnel only, and all records of visits are to be maintained.
    (*h*) Full account of transgenic seeds produced are to be kept and no part of this seed can be transacted or future propagated without authorization.

38.11. GUIDELINES AND REGULATIONS

The guidelines for recombinant DNA research and for planned introduction of GMOs should be such that they encourage such research rather than impede them and, at the same time minimise the hazards from such experiments. The guidelines should set conditions for laboratory research as well as small scale field research; there should be minimum of regulation but reporting must be adequate. The conditions for research should be designed according to the scientific and transparent risk assessment; they should not be based on *a priori* high risk categorization of all research using recombinant DNA technology. Stringent conditions for preliminary research could hamper the development of biotechnology industry. Therefore, it has been suggested that academic and preliminary phases of private research should be supervised by both public and government agencies, but regulatory authorities should move in only after a potential product has been identified and is likely to be marketed; therefore an all encompassing supervision system should be developed.

The safety and purpose of the proposed research should not be evaluated solely on considerations of potential economic benefits perceived at the time. The restrictions should be based on demonstrated rather than imagined risk. The information required for risk assessment should not be such that the cost of generating such information effectively prohibits the experiments. There should be effective coordination among different government agencies involved in the regulation of research and products, and the responsibilities and authorities of each agency should be clearly specified and commensurate with its traditional responsibilities and authorities.

A uniform set of principles of safety should be used by all research workers regardless of the source of funding and the type of institution (private or public institution) where the research work is being conducted. The guidelines should be prepared in accordance with not only scientific and social/economic considerations, but also from the point of view of legal validity. It is more than likely that many decisions of regulatory authorities would be challenged in courts of law, where they will have to be defended.

The guidelines and regulations governing recombinant DNA research activities vary considerably among countries and many countries, especially third world countries, do not have such guidelines and regulations. It is highly desirable to harmonize the guidelines and regulations of various countries otherwise research workers and or organisations, especially private companies, may shift their research experiments to such countries that have more lax regulations or no regulations at all. In the past, several such shifts (***relocation of experiments***) were made, some of which became scandals.

In 1991, the Informal Working Group on Biosafety (IWGB), the main body of UN on biosafety, issued the draft voluntary code of conduct for Release of Organisms into the Environment. This document covers biosafety aspects related to public health and the environment at all the stages of research, development, use and disposal of GMOs. This code *has not* been designed as a manual for handling of GMOs. The IWGB recommends the establishment of an International Biosafety Information Network and Advisory Service, which would function as a steering committee and will set up advisory panels on request.

The Consultative Group on International Agricultural Research (CGIAR) created a taskforce on biotechnology (BIOTASK) in 1989. BIOTASK has on its agenda, along with other things, regulatory issues and environmental release of GMOs related to agriculture.

QUESTIONS

1. Explain the concept of risk assessment relevant to recombinant DNA technology with special reference to laboratory experiments (or planned introduction).
2. Briefly explain the concept of containment during recombinant DNA research.
3. Briefly explain the concepts of physical and biological containment during recombinant DNA research.
4. Discuss the various issues related to planned introduction of genetically modified microorganisms (or plants).
5. Explain the objectives of guidelines for recombinant DNA technology, and discuss the role of regulations in the development of biotechnology industry.

6. Write short notes on the following:
   (*i*) genetically modified organisms,
   (*ii*) planned introduction,
   (*iii*) containment,
   (*iv*) risk assessment,
   (*v*) physical containment,
   (*vi*) biological containment,
   (*vii*) biosafety,
   (*viii*) host-vector systems,
   (*ix*) risk groups,
   (*x*) initial risk assessment,
   (*xi*) comprehensive risk assessment,
   (*xii*) risk assessment for planned introduction,
   (*xiii*) risk assessment of biotechnology products,
   (*xiv*) risk regulation,
   (*xv*) biosafety levels.

39

Cytoplasmic Inheritance and Maternal Effects

39.1. MENDELIAN INHERITANCE

Inheritance of most of the characters of eukaryotic organisms shows the following characteristic features: (1) contribution by both male and female parents is equal so that results from reciprocal crosses are identical; and (2) segregation produces the characteristic 3 : 1 (or 1 : 2 : 1) ratio in the F_2 generations of monohybrid crosses and a typical 9 : 3 : 3 : 1 (or one of its modifications in the cases of gene interaction) in dihybrid crosses. These features of inheritance were first demonstrated by Mendel. Consequently, such an inheritance pattern is referred to as ***Mendelian inheritance.*** *It is universally accepted that genes showing Mendelian inheritance are located in chromosomes of eukaryotic nuclei.* Therefore, Mendelian inheritance pattern is regarded as a sufficient evidence for a gene to be located in chromosomes. Such genes are termed as ***nuclear genes*** or, more commonly, simply as ***genes.***

39.2. NONMENDELIAN INHERITANCE

But some characters in several organisms do not show Mendelian inheritance. In such cases, the following characteristic features are found. (1) There is consistent difference between results from reciprocal crosses. (2) Generally, trait from the female parent only is transmitted. (3) In most cases, there is no segregation in F_2 and subsequent generations.

Nonmendelian inheritances may be grouped under three broad categories : (1) those related to cellular structures and patterns, (2) those produced by intracellular parasites, symbionts and viruses, and (3) those associated with DNA containing cell organelles, *viz.,* mitochondria and chloroplasts.

39.3. HISTORICAL

Evidence for cytoplasmic inheritance was first presented by Correns in *Mirabilis jalapa* and by Baur in *Pelargonium zonale* in 1908, and they termed this inheritance pattern as ***plastid inheritance.*** In 1924, Jenkins described the *iojap* leaf variegation in maize, which was later investigated in considerable detail by Rhoades.

Rhoades described cytoplasmic male sterility in maize in 1933. Ten years later, in 1943, Sonneborn discovered *kappa* particles in *Paramecium*, and described their cytoplasmic transmission. In 1954, Ruth Sager described cytoplasmic (uniparental) inheritance of streptomycin resistance in *Chlamydomonas.* Presence of DNA in chloroplasts was first demonstrated by Ris and Plaut in 1962. One year later, in 1963, Nass and coworkers proved the existence of DNA in mitochondria.

39.4. DEFINITIONS

In cases of ***cytoplasmic inheritance,*** generally, the character of only one of the two parents (usually the female parent) is transmitted to the progeny. As a result, reciprocal crosses exhibit consistent differences for such characters, and there is a lack of segregation in F_2 and subsequent generations. Such inheritance is also referred to as ***extranuclear inheritance, extrachromosomal inheritance*** and ***maternal inheritance.*** Since cytoplasmic inheritance is more commonly used, it is preferred to the other terms in this text.

Since the genes governing traits showing cytoplasmic inheritance are located outside the nucleus and in the cytoplasm, they are referred to as ***plasmagenes, cytoplasmic genes, cytogenes, extranuclear genes*** or ***extrachromosomal genes***. Out of these terms, plasmagene has been used in the present discussion.

The sum total of genes present in the cytoplasm of a cell or an individual is known as ***plasmon,*** while all the genes in a plastid constitute a ***plastome***. Similarly, all the genes present in a mitochondrion together constitute the ***chondriome***. All the available evidence indicates that plasmagenes are located in DNA present in mitochondria (abbreviated as mtDNA) and chloroplasts (shortened as cpDNA).

Maternal Effects

- Nuclear genotype of mother affects character development.
- Reciprocal differences in F_1.
- In extreme case, *e.g.*, coiling in snails:
  - (*i*) No observable segregation in F_2.
  - (*ii*) Segregation (3 : 1) observable in F_3.
- The concerned traits obviously governed by nuclear genes.

39.5. MATERNAL EFFECTS

The development of some characters in several organisms is either governed or markedly influenced by the genotype of the female parent; this is known as ***maternal effect***. Such characters, are governed by nuclear genes. But due to the maternal effect, they show the following features: (1) reciprocal differences in F_1; (2) which, in most cases, disappear in F_2; and (3) a considerably smaller variation in F_2 as compared to that in F_3. In some extreme cases, there may be no phenotypic segregation in F_2. In such cases, phenotypic segregation is evident only in the F_3 generation. Many important characteristics of both plants and animals show maternal effects, of which some examples are briefly described below.

39.5.1. Coiling in *Limnaea*

An extreme example of maternal effect is known in the snail *Limnaea.* In this snail, the direction of coiling of its shell is controlled by a single nuclear gene *D/d.* The dominant allele *D* produces right-handed (= dextral) coiling, while its recessive allele *d* produces left-handed (= sinistral) coiling. ***The direction of shell coiling in an individual is governed by the genotype of its female parent and not by is own genotype.*** *As a result, reciprocal crosses show differences in coiling in* F_1 *and there is no phenotypic segregation in* F_2. *The phenotypic effect of segregation is observable in* F_3 *only.*

Crosses between females with left-handed coil (*dd*) and males having right-handed coil (*DD*) produce F_1 progeny (*Dd*) with left-handed coil (since the genotype of female parent is *dd*). Segregation for *Dd* (in F_1) produces three genotypes, *DD, Dd,* and *dd,* in the F_2, the ratio being 1 : 2 : 1. But the snails with *DD, Dd* as well as *dd* genotypes exhibit right-handed coiling since their female parents had the genotype *Dd*. The F_3 progeny from F_2 female individuals with the genotypes *DD* and *Dd* will show right-handed coiling, while those from *dd* F_2 female individuals will exhibit left-handed coiling. This produces the typical 3 : 1 ratio (right-handed : left-handed) in F_3 (Fig. 39.1).

The reciprocal cross (female right-handed, *DD* × male left-handed, *dd*), on the other hand, yields right-handed coiling in F_1 (*Dd*) as well as in the three genotypes (1 *DD* : 2 *Dd* : 1 *dd*) obtained in F_2. But in F_3, 3/4 of the progenies show right-handed coiling since they are derived from F_2 females with genotypes *DD* and *Dd*. The remaining **1/4** of the F_3 progeny exhibit left-handed coiling since their female parents had the genotype *dd;* this yields the typical monohybrid ratio of 3 : 1 in F_3.

The features of inheritance of coiling in *Limnaea* shells may be summarised as follows : (1) F_1's from reciprocal crosses show differences in coiling pattern, (2) there is homogeneous coiling in F_2, *i.e.,* there is a lack of observable segregation in F_2, and there is (3) the typical 3 : 1 ratio is exhibited in F_3 in place of F_2. *The* 3 : 1 *ratio (although observed in* F_3*) clearly indicates that coiling of shells is governed by a single nuclear gene.* But the segregation of this nuclear gene appears to be delayed by one generation. In fact, segregation of the nuclear gene, *Dd,* is normal and occurs during the gamete formation in the F_1 generation itself. But the phenotypic effects of this segregation become visible only one generation later, *i.e.,* in F_3, than the usual, *i.e.,* in F_2.

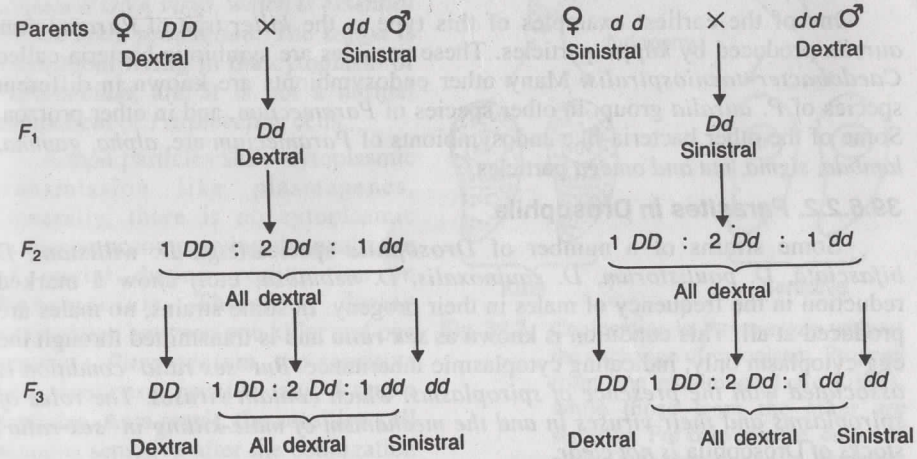


Fig. 39.1. Maternal effect on the direction of shell coiling in *Limnaea* (a snail), and the delayed segregation for coiling. The coiling of an individual is determined by the genotype of its maternal parent.

39.5.2. Characters of Plants

The development of many traits in higher plants, including many crop species, is affected by the genotype of the female parent. Examples of such characters include such economically important traits as seed size, protein content in seeds, and, in some species, plant height, pod size, etc. Consequently, F_1 progeny from reciprocal crosses show marked differences for such characters. These differences, however, disappear in F_2, and the amount of variation present in F_2 is usually smaller than that observed in F_3.

39.6. NONMENDELIAN INHERITANCE

As stated earlier, there are three distinct classes of nonmendelian inheritance : (1) inheritance of structures and patterns, (2) inheritance due to parasites, symbionts and viruses, and (3) cytoplasmic inheritance.

39.6.1. Inheritance of Structures and Patterns

It was once believed that certain cellular structures and organizational patterns are transmitted from one generation to the next and show inheritance independent of nucleus. Examples of such structures are bacterial cell wall, cellular membranes and centrioles, and the pattern of arrangement of cortical structures (ciliary basal bodies, trichocysts, fibers and microtubules) in *Paramecium*. But *it is now definitely known that these patterns and structures are governed by nuclear genes only.*

39.6.2. Nonmendelian Inheritance Due to Parasites, Symbionts and Viruses

A number of cases were once regarded as classical cases of cytoplasmic inheritance, but they are actually due to the presence of parasites, symbionts or viruses. Some such examples are briefly described below.

39.6.2.1. Symbionts in Paramecium *and Other Protozoa*

One of the earliest examples of this type is the *killer* trait of *Paramecium aurelia* produced by *kappa* particles. These particles are symbiotic bacteria called *Caedobacter taeniospiralis*. Many other endosymbionts are known in different species of *P. auralia* group, in other species of *Paramecium*, and in other protzoa. Some of the other bacteria-like endosymbionts of *Paramecium* are, *alpha, gamma, lambda, sigma, mu and omega* particles.

39.6.2.2. Parasites in Drosophila

Some strains of a number of *Drosophila* species (*e.g., D. willistoni, D. bifasciata, D. paulistorum, D. equinoxalis, D. nebulosa,* etc.) show a marked reduction in the frequency of males in their progeny. In some strains, no males are produced at all. This condition is known as ***sex ratio*** and is transmitted through the egg cytoplasm only, indicating cytoplasmic inheritance. *But 'sex ratio' condition is associated with the presence of spiroplasms, which contain viruses. The roles of spiroplasms and their viruses in and the mechanism of male-killing in 'sex-ratio' stocks of* Drosophila *is not clear.*

39.6.2.3. Viruses in Drosophila

Some strains of *D. melanogaster* are sensitive to CO_2 as they die when briefly exposed to CO_2, while normal flies are not affected by a similar treatment.

CO_2-sensitivity is associated with the presence of a virus-like *sigma* factor. ***Sigma factor*** is transmitted through the egg cytoplasm so that progeny from the cross CO_2-sensitive female × normal male are all CO_2-sensitive, while those from the reciprocal cross (normal female × CO_2-sensitive male) are all normal. CO_2-sensitive stocks are known in some other species of *Drosophila. Electron microscope studies have clearly shown the sigma factor to be a virus, resembling in shape the 'torpedo' stage of developing plant embryos.*

Similar apparent cases of cytoplasmic inheritance produced by viruses, etc. are known in several other organisms like fungi, mice, etc.

39.6.2.4. Kappa Particles of Paramecium

Some strains of *Paramecium* contain *kappa* particles (*Caedobacter taeniosiralis*) in their cytoplasm. *Kappa* particles are called either (1) bright or (2) nonbright, depending on their appearance under the light microscope. *Bright kappa particles are believed to produce a toxin, paramecin, which is lethal to sensitive strains of* Paramecium. All *Paramecium* cells lacking kappa are ***sensitive*** to paramecin. *Paramecium* containing kappa particles are called ***killers***. They produce the toxin, and are resistant to the toxin. *Bright kappas contain a DNA virus, which is essential for the toxin production.* The *kappa* is a symbiont living in the cytoplasm of *Paramecium,* and it is not a normal component of *Paramecium* cells.

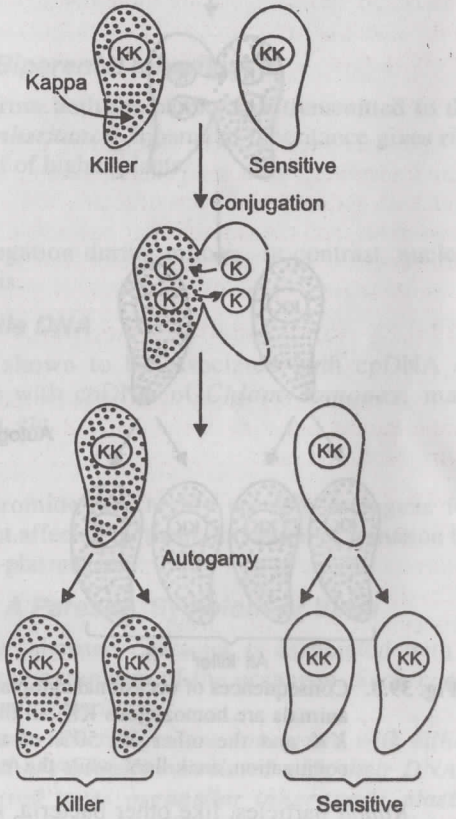


Fig. 39.2. Conjugation in *Paramecuim* without the exchange of cytoplasm. Progeny from one exconjugant are killers, while those from the other are sensitive. For convenience only one of the two micronucli is depicted, and the macronucleus is not shown.

Kappa particles show cytoplasmic transmission like plasmagenes. Generally, there is no cytoplasmic exchange accompanying the exchange of nuclei during conjugation in *Paramecium*. Therefore, during conjugation between one killer and one sensitive *Paramecium*, the sensitive cell does not receive any *kappa* particles. As a result, the sensitive cell remains sensitive after the conjugation and produces sensitive progeny (Fig. 39.2).

But occasionally, cytoplasmic exchange does occur during conjugation; in such cases, both the animals that participated in conjugation (exconjgants) receive *kappa.*

But subsequent propagation of *kappa* depends on the nuclear genotypes of these animals (Fig. 39.3).

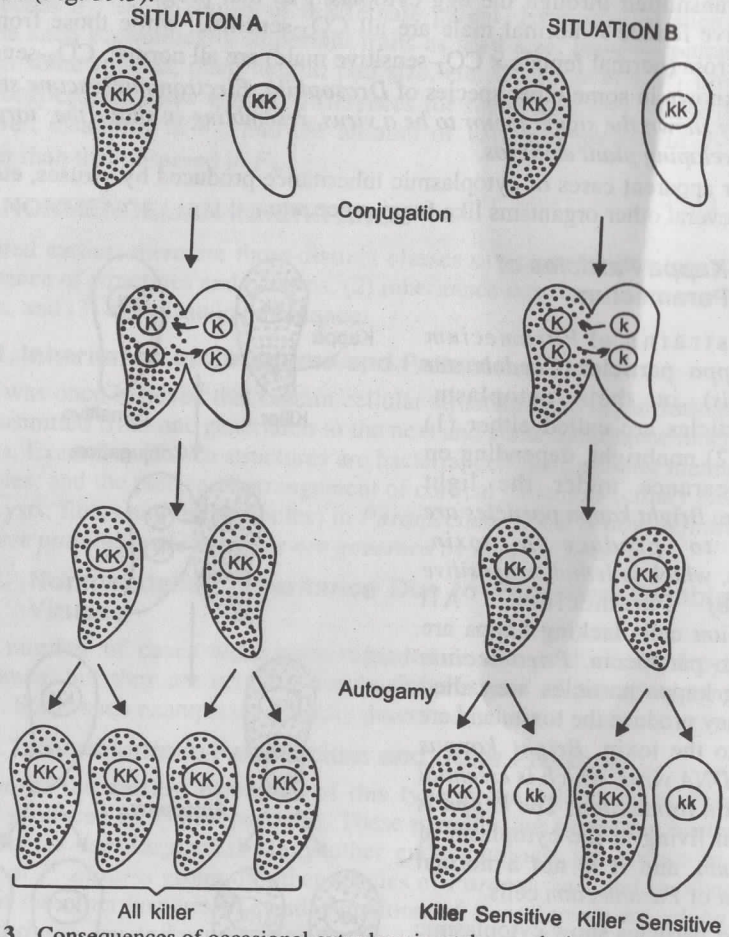


Fig. 39.3. Consequences of occasional cytoplasmic exchange during conjugation. When both animals are homozygous KK, all the progeny are killers. But when one animal is *KK* and the other *kk*, 50% of the progeny following autogamy, after the conjugation, are killers, while the rest (50%) are sensitives.

Kappa particles, like other bacteria, multiply through fission. However *kappa particles are able to multiply only in animals having the genotypes* KK *or* Kk. *In* kk *animals,* kappa *are unable to multiply.* As a consequence, they are eliminated from such animals. Thus killer animals with *kk* genotype ultimately become sensitive due to a loss of their *kappa* particles

39.6.3. Cytoplasmic Inheritance

The foregoing discussion was focussed on traits due to cellular structures and patterns, and on those due to parasites, endosymbionts and viruses. ***These cases, do not represent cytoplasmic inheritance.*** *Cytoplasmic inheritance is due to plasmagenes located in cell organelles (plastids and mitochondria) that are integral*

constituents of normal cells. The characteristic features of this type of inheritance are summarised below.

39.6.3.1. Reciprocal Differences

Reciprocal crosses show marked differences for the characters governed by plasmagenes. In most cases, plasmagenes from only one parent, generally the female parent, are transmitted. This phenomenon is known as ***uniparental inheritance.***

39.6.3.2. Lack of Segregation

In general, F_2, F_3 and the subsequent generations do not show segregation for a cytoplasmically inherited trait.

39.6.3.3. Irregular Segregation in Biparental Inheritance

In some rare cases, plasmagenes from both the parents are transmitted to the progeny. This is known as ***biparental inheritance.*** Biparental inheritance gives rise to irregular segregation ratios in F_1 itself of higher plants.

39.6.3.4. Somatic Segregation

Plasmagenes generally show segregation during mitosis. In contrast, nuclear genes ordinarily segregate during meiosis.

39.6.3.5. Association with Organelle DNA

Several plasmagenes have been shown to be associated with cpDNA or mtDNA, *e.g.,* streptomycin resistance with cpDNA of *Chlamydomonas,* male sterility with mtDNA in maize, sorghum, etc.

39.6.3.6. Mutagenesis

Some mutagens, *e.g.,* ethidium bromide, are highly specific mutagens for plasmagenes, while nuclear genes are not affected by them. Induction of mutation by such agents in a gene indicates it to be a plasmagene.

39.6.3.7. Lack of Association with A Parasite, Symbiont or Virus

In many cases, a cytoplasmically inherited character is associated with a parasite, symbiont or virus present in the cytoplasm of the organism. *Such cases cannot be regarded as cases of cytoplasmic inheritance.*

The known cases of true cytoplasmic inheritance are concerned with either chloroplast or mitochondrial traits and are usually associated with their DNAs. Such cases are, therefore, often referred to as ***organellar inheritance, plastid inheritance*** or ***mitochondrial inheritance.*** A more detailed account of some of these follows.

39.7. PLASTID INHERITANCE

Inheritance of plastid characters due to plasmagenes located in plastids is known as ***plastid inheritance.*** All our knowledge about plastid inheritance in higher plants is based on studies on chlorophyll variegation in leaves. ***Variegation*** refers to the presence of white or yellow spots of variable size on the green background of leaves. Mutant plasmagenes affecting chloroplasts are known in many crop plants, *e.g.,* barley, wheat, jowar, maize, pea, tobacco, tomato, many ornamental plants, etc.

Plastid Inheritance

- Due to Plastome.
- Usually, chlorophyll variegation in leaves.
- Also streptomycin and atrazine resistances.
- *First report*: Correns (1908) and Baur (1908).
- Typical cytoplasmic inheritance.
- When variegated branches used as female, green, white and variegated progeny produced.
- Maternal transmission due to sperm contributing virtually no cytoplasm to zygote.

39.7.1. Mirabilis jalapa

Leaves of *Mirabilis jalapa*, the four O'clock plant, may be green, white, or variegated. Some branches may have only green, only white or only variegated leaves. Correns made reciprocal crosses in all combinations among the flowers produced on these three types of branches; results from these crosses are summarised in Table 39.1. When flowers from a green branch were used as female, all the progeny were green irrespective of the phenotype (green, white or variegated) of the male parent. Similarly, progeny from the crosses involving flowers from white branches as the female parent were all white irrespective of the male flowers being from green, white or variegated branches. But in the progeny from the different crosses involving flowers from variegated branches as the female parent, green, white and variegated plants were recovered in variable proportions.

TABLE 39.1
A summary of the findings of Correns from various types of crosses in *Mirabilis jalapa*

Leaf phenotype of the branch used as female parent	*Leaf phenotype of the branch used as male parent*	*Leaf phenotype of progeny* (F_1)
Green	1. Green	Green
	2. White	Green
	3. Variegated	Green
White	1. Green	White
	2. White	White
	3. Variegated	White
Variegated	1. Green	Green, white and variegated in variable ratios in each of the cases
	2. White	
	3. Variegaged	

Note : Leaf phenotype of the progeny is the same that of the female parent. In the progeny from variegated female parents, green, white and variegated progeny are recovered in variable proportions.

Clearly, *the phenotype of progeny was the same as that of the female parent, except when the female parent was variegated.* These results may be explained by assuming that the plasmagene producing variegation in *M. jalapa* is located in plastids. Chloroplasts differentiate from proplastids. Proplastids having the mutant

plasmagene differentiate into colourless abnormal chloroplasts, while those having the normal allele develop into normal green chloroplasts.

The *maternal transmission of plasmagenes in higher plants and animals is explained on the basis of unequal contribution by male and female gametes to the cytoplasm of zygote.* In general, egg/egg cell has a considerable amount of cytoplasm, which contains several mitochondria and proplastids. On the other hand, male gametes or sperms have only a thin layer of cytoplasm surrounding the nuclei, which usually does not contain any mitochondira or proplastids. Further, during fertilization, ordinarily only the sperm nucleus enters egg cell, and its cytoplasm is left outside the egg cell. As a result, the mitochondria and proplastids present in the zygote are derived from the female parent only (Fig. 39.4). Therefore, plasmagenes located in these organelles will also be derived from the female parent only.

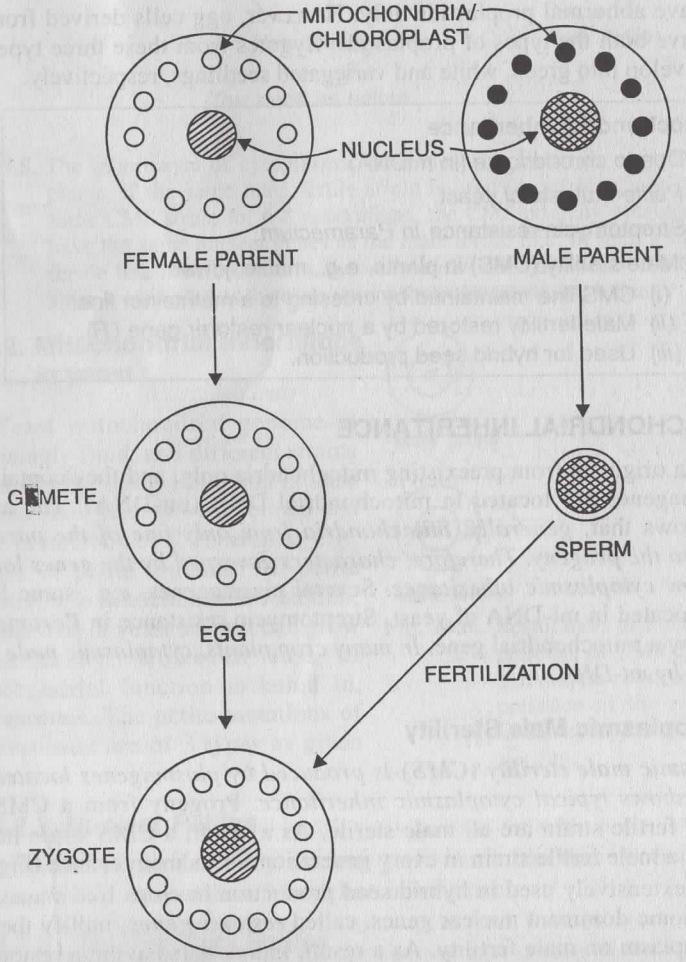


Fig. 39.4. Unequal contribution of the cytoplasm by the male and female gametes as the basis of cytoplasmic inheritance in higher plants and animals.

Now we return to the results from various crosses in four O'clock. Green leaves obviously have only normal chloroplasts, while white leaves have only abnormal ones. In variegated leaves, green sectors will have normal plastids, while white sectors will have abnormal ones. Egg cells produced on a green branch will have normal proplastids, while sperms produced on green, white or variegated branches will contribute no plastids to the zygote. Therefore, progeny from the three crosses using flowers from green branches as females will all be green. Similarly, progeny from the three crosses involving flowers on white branches as female will all be white since zygotes from these crosses will receive only the abnormal proplastids through egg cells. However, cells in branches having variegated leaves may have only normal (corresponding to the green sector), only abnormal (corresponding to the white sector) or both the types (mixed cells) of proplastids. Egg cells from 'only normal' cells will have only normal proplastids, while those from 'only abnormal' cells will have abnormal proplastids only. However, egg cells derived from mixed cells will have both the types of proplastids. Zygotes from these three types of egg cells will develop into green, white and variegated seedlings, respectively.

Mitochondrial Inheritance

- Due to chrondriome (in mtDNA)
- *Petite* mutants of yeast.
- Streptomycin resistance *in Paramecium.*
- Male sterility (CMS) in plants, *e.g.,* maize, jowar.
  - (*i*) CMS line maintained by crossing to a maintainer line.
  - (*ii*) Male fertility restored by a nuclear restorer gene (*R*).
  - (*iii*) Used for hybrid seed production.

39.8. MITOCHONDRIAL INHERITANCE

Mitochondria originate from preexisting mitochondria only, and they contain DNA. Many plasmagenes are located in mitochondrial DNA (mt-DNA). The available evidence shows that, *generally, mitochondria from only one of the parents are transmitted to the progeny. Therefore, characters governed by the genes located in mt-DNA show cytoplasmic inheritance.* Several plasmagenes, *e.g.,* some kinds of *petites,* are located in mt-DNA of yeast. Streptomycin resistance in *Paramecium* is determined by a mitochondrial gene. *In many crop plants, cytoplasmic male sterility is controlled by mt-DNA.*

39.8.1. Cytoplasmic Male Sterility

***Cytoplasmic male sterility* (*CMS*)** *is produced by plasmagenes located in mt-DNA. CMS shows typical cytoplasmic inheritance.* Progeny from a CMS and a normal male fertile strain are all male sterile. As a result, a CMS strain has to be pollinated by a male fertile strain in every generation for its maintenance (Fig. 39.5).

CMS is extensively used in hybrid seed production in crops like maize, jowar, bajara, etc. Some dominant nuclear genes, called ***restorer genes***, nullify the effects of CMS cytoplasm on male fertility. As a result, individuals having a restorer gene in homozygous or heterozygous state are male fertile even in the presence of CMS cytoplasm (Fig. 39.6).

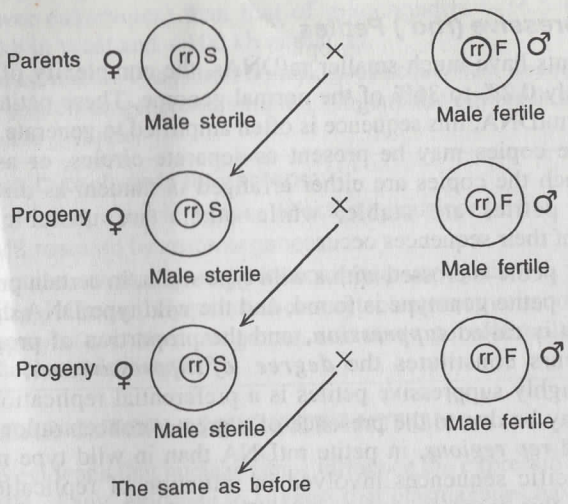


Fig. 39.5. The inheritance of cytoplasmic male sterility (CMS) in maize and other higher plants. If the same male fertile strain is used in crosses (back crosses) with the same CMS strain for 6-8 generations, the CMS progeny obtained in the end will have the same nuclear genes as the male fertile line. In such a situation, the male fertile line is referred to as the *maintainer line* (*B line*) of the *CMS line*(*A line*) since it is needed as the male parent for maintaining the A line.

39.8.2. Mitochondrial Inheritance in yeast

Yeast mitochondrial genome is surprisingly fluid, and different strains may differ by upto 10,000 bp in their genome size. Extensive deletions can occur in mtDNA of *S. cerevisiae, e.g.,* in case of petite mutants. All petite mutants lack mitochondrial function; they survive in yeast since it can grow anaerobically. However, loss of mitochondrial function is lethal in, say, animals. The petite mutations of *S. cerevisiae* are of 3 types as given below.

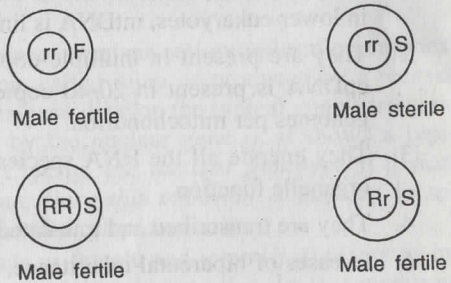


Fig. 39.6. Restoration of male fertility in the presence of CMS cytoplasm (*S*) by a dominant restorer gene, *R*. In the presence of the recessive alle, *r*, CMS cytoplasm produces male sterility.

39.8.2.1. Nuclear Petites

These are produced by nuclear gene mutations that abolish mitochondrial function. These petites show Mendelian inheritance.

39.8.2.2. Neutral (rho^0) Petites

These petites are characterised by the absence of mtDNA; they represent an extreme situation and are recessive in inheritance.

39.8.2.3. Suppressive (rho⁻) Petites

These mutants have much smaller mtDNAs, the complexity of their genomes ranging from only 0.2% to 36% of the normal genome. These petites may contain any sequence of mtDNA; this sequence is often amplified to generate a large number of copies. These copies may be present as separate circles, or as a multimeric molecule in which the copies are either arranged in tandem as direct or inverted repeats. Some petites are stable, while others are unstable, and further rearrangements of their sequences occur.

When a *rho*⁻ petite is crossed with a wild type strain, in certain proportion of the progeny only the petite genotype is found, and the wild type DNA disappears from them. This effect is called ***suppression,*** and the proportion of progeny in which suppression occurs constitutes the ***degree of suppressiveness.*** The cause of suppression in highly suppressive petites is a preferential replication of the petite genome. This may be due to the presence of a greater concentration of origin-like sequences, called *rep regions,* in petite mtDNA than in wild type mtDNA. Yeast mtDNA has specific sequences involved in initiation of replication. But petite mtDNAs lacking this sequence can survive. This indicates that all the segments of mtDNA in yeast can be independently replicated since any sequence of mtDNA can be present in a petite. The molecular basis for the phenomenon of suppression is, however, unknown.

39.9. FEATURES OF ORGANELLAR GENOMES

1. They are circular molecules of DNA. In few exceptional cases, generally in lower eukaryotes, mtDNA is linear.
2. They are present in multiple copies in each organelle. In higher plants, cpDNA is present in 20-40 copies per chloroplast. Yeast cells have ~4 genomes per mitochondrion.
3. They encode all the RNA species and some of the proteins required for organelle function.
4. They are transcribed and translated within the organelles.
5. In cases of biparental inheritance, recombination does take place between organellar genomes.
6. Organelle DNA is replicated by a different DNA polymerase from that of the nucleus.
7. The DNA repair systems and other events that impinge on the fidelity of DNA sequences are different in organellels from those in nuclei.
8. Accumulation of mutation is much faster in mtDNA than in nuclear genome in the case of mammals, while in plants the reverse is the case. The cpDNA is intermediate between mtDNA and nuclear DNA. Mammalian mtDNA accumulates mutations at 2-4% per million years, which is >10 × faster than the rate for globin genes.
9. In case of uniparental inheritance in *Chlamydomonas reinhardii,* the cpDNA contributed by one parent is degraded shortly after zygote formation, leaving in the zygote cpDNA from only one of the parents.

10. Chloroplast genome is much larger (~140 kb in higher plants and <200 kb in lower eukaryotes) than that of mitochondria (~16.5 kb in mammals, ~ 84 kb in yeast and ~ 100 kb in plants).
11. In plants, mtDNA has short homologous sequences. Recombination within these sequences generates smaller, subgenomic circles that coexist with the complete genome.

Nuclear-Cytoplasmic Interactopm

- Expression of plasmagenes affected by nuclear genes.
- CMS restored by *restorer* genes.
- In maize, nuclear gene *ij* produces *iojap* variegation.
- Once produced, iojap trait shows cytoplasmic inheritance.
  Explanation: Nuclear genes contribute to chloroplast and mitochondrial function.

39.10. NUCLEAR-CYTOPLASMIC INTERACTION

There is ample evidence that nuclear genes influence the expression of plasmagenes. But there is little, if any, evidence to suggest that plasmagenes have a similar effect on the latter. The expression of plasmagenes is, as a rule, associated with either mitochondrial or plastid function. Nuclear genes contribute significantly to both mitochondrial and chloroplast functions as well as structures. It is, therefore, not surprising that the expression of their structure and function is greatly modified by nuclear genes. A classical example of interaction between nuclear and cytoplasmic genotypes is that between CMS and restorer genes (Section 39.8.1).

39.10.1. Iojap Variegation of Maize

An unusual case of interaction between nuclear and cytoplasmic genomes is reported in maize. A type of variegation, called iojap, is produced by a recessive gene *ij*. Plants homozygous (*ij ij*) for this gene develop the typical iojap variegation. But once this variegation is produced by the nuclear gene *ij,* it shows a typical cytoplasmic inheritance (Fig. 39.7). Clearly, *the nuclear genotype* ij ij *has a mutagenic effect on the plastid genome. Once this mutation is induced in some cpDNA molecules, the variegation is inherited cytoplasmically.*

A cross between normal (I_j I_j) plants as female and iojap (*ij ij*) plants as male produces all green plants in F_1 with the nuclear genotype *Ij ij*. In F_2 generation of this cross, 1/4 progeny are *ij ij*; these plants develop the iojap variegation. The remaining 3/4 of the progeny are normal green (Fig. 39.7). When the iojap F_2 plants are mated with normal green plants, a marked reciprocal difference is observed in the progeny. When iojap plants are used as male and green plants as female, all the progeny are normal green. But in the reciprocal cross, green, white and iojap progeny are recovered; the ratio of the three types of progeny is quite variable. Obviously, the iojap trait, once produced, shows cytoplasmic or plastid inheritance.

39.10.2. Molecular Basis of Nuclear—Cytoplasmic Interaction

Most of the proteins needed for mitochondrial and chloroplast function are imported from the cytoplasm and encoded by nuclear genes. For example, mtDNA is estimated to encode only 10% of the proteins needed for mitochondrial structure

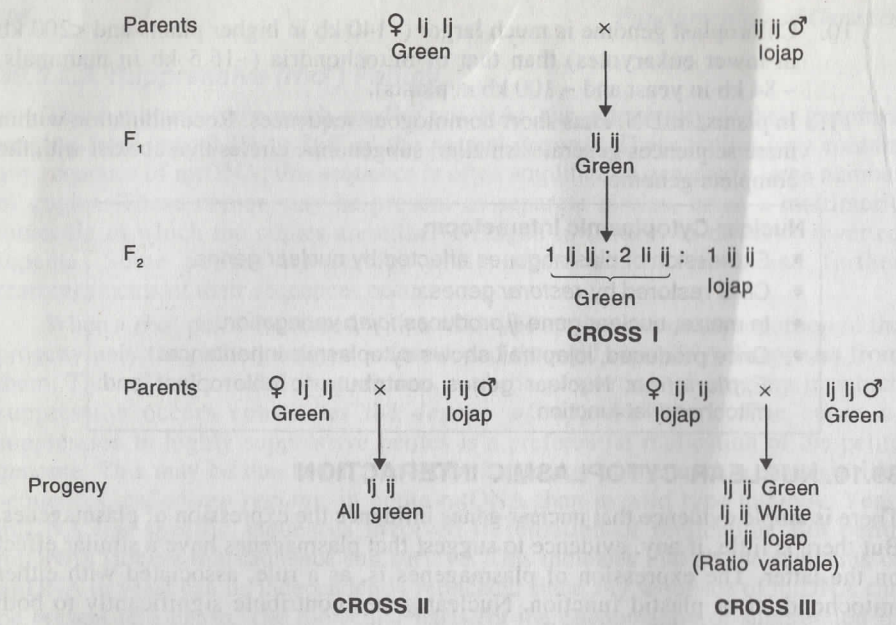


Fig. 39.7. The production of iojap variegation in maize by a recessive nuclear gene *ij* (Cross *I*). But once the variegation is produced, it shows cytoplasmic inheritance (Crosses *II* and *III*).

and function; each of these proteins forms a component of a multimeric protein molecule that includes some subunits encoded by nuclear genes. The organelles depend on the nuclear genome in one of the following three ways.

1. The functional protein molecules consist of some subunits encoded by organelle genes, while the rest are encoded by nuclear genes. For example, mitochondrial ATPase (oligomycin sensitive ATPase) has two subunits, *viz.*, F0 membrane factor and F1 ATPase. F0 has 3 polypeptides, which are encoded by mtDNA in yeast. The F1 ATPase has 5 subunits all of which are coded for by nuclear genes in yeast. There is, however, some variation among species in the contributions to ATPase by the nuclear and mitochondrial genomes (Table 39.2).

TABLE 39.2

The contribution of organellar and nuclear genomes to organelle ATPase. The notations within parenthesis represent the polypeptides of ATPase

Organism	*F_0 Factor*		*F_1 ATPase*	
	Organelle	*Nucleus*	*Organelle*	*Nucleus*
S. cerevisiae (mtDNA)	3 (6, 8, 9)	0	0	5 (1, 2, 3, 4, 7)
A. nidulans (mtDNA)	3 (6, 8, 9)	0	0	5 (1, 2, 3, 4, 7)
Neurospora Crassa (mtDNA)	2 (6, 8)	1 (9)	0	5 (1, 2, 3, 4, 7)
Plants (cpDNA)	2 (I, II)	1 (II)	2 (γ, δ)	4 (α, β, ε, a)

2. The protein molecules are encoded solely by nuclear genes; most of the organelle proteins fall under this category. For example, most or all of the protein components of chloroplast and mitochondrial protein synthesis are of nuclear origin; these include ~ 75 ribosomal proteins and 22 amino acyl-tRNA synthetases. However, all of the protein components of the synthetic apparatus seem to be unique to the mitochondrion, *i.e.*, they are encoded by the nuclear genes especially for mitochondria and are not used in protein synthesis in the cytoplasm.
3. Some nuclear genes encode products that are necessary for the assembly of the oligomers from polypeptides encoded by mitochondrial and nuclear genes. These polypeptides, however, do not form a part of the final oligomeric aggregate (*e.g.*, ATPase). The assembly of each of these aggregates appears to be a process in which the absence of one component prevents the assembly of the others.

Obviously, a proper organellar function would depend on many nuclear genes and this would lead to a strong interaction between the nuclear genes and plasmagenes.

QUESTIONS

1. Explain the various characteristic features of cytoplasmic inheritance with the help of suitable examples.
2. Explain mitochondrial or plastid inheritance with the help of a suitable example, and explain nuclear-cytoplasmic interactions.
3. 'Characters showing cytoplasmic inheritance in many organisms are produced by parasites, symbionts or viruses.' Critically analyse this statement.
4. Write short notes on the following:
   (*i*) maternal effect, (*ii*) cytoplasmic inheritance,
   (*iii*) symbionts, (*iv*) plastid inheritance,
   (*v*) mitochondrial inheritance, (*vi*) nuclear-cytoplasmic interaction.

40

Population Genetics
I. The Hardy-Weinberg Law

40.1. INTRODUCTION

The pattern of inheritance of qualitative traits is determined through the technique of classical genetic analysis originally developed by Mendel, which is based on controlled matings. For example, the dominant homozygote *AA* is crossed with the recessive homozygote *aa* and the F_1, *Aa*, is either selfed to yield F_2, or it is crossed to the recessive homozygote (*aa*) to generate test-cross progeny. In contrast, natural populations are generally random mating. In ***random mating***, every male gamete produced by the population has equal opportunity or probability of uniting with any female gamete of the population, and *vice-versa*. In other words, each male of a population has equal chance of mating with any female of the population, and *vice-versa*. Such populations are called ***Mendelian populations*** (first used by Wright), ***random mating populations*** or ***panmictic populations.***

In random mating populations, (1) *there is no controlled mating,* (2) *the pattern of transmission of individual genes cannot be determined,* (3) *the genotypes of one or few individuals are not important, and* (4) *gene pool and gene frequencies are the two main attributes. As all the individuals of a Mendelian population intermate freely, they must belong to the same species, more particularly, the same race, and must occupy the same territory or area since matings among individuals located at a distance are expected to be infrequent.* Thus each species may be considered to be made up of several Mendelian populations, each occupying a separate territory; these populations are often called ***demes*** and they are the basis of race formation (see, Chapter 42). Further, by the virtue of random mating, all the individuals of a Mendelian population share the same gene pool (see later). Therefore, ***population genetics*** *aims at understanding the dynamics of gene and genotype frequencies in random mating populations.*

40.2. GENE FREQUENCY

The proportions of different alleles of a gene present in a Mendelian population are known as ***gene frequency.*** Thus the proportions of gametes produced by a population carrying the different alleles of a gene are the frequencies of these alleles in the population, *i.e.*, gene frequency. Gene frequencies in a population can be readily estimated by classifying the individuals of a random sample from the population into different genotypic classes for a gene. The total number of each of the alleles of the gene present in these individuals is then computed, and their ratio to the total number for all the alleles of the gene is estimated; these ratios are the gene frequencies for that locus (Table 40.1).

TABLE 40.1

Estimation of the frequencies of L^M and L^N alleles of the gene producing MN blood group in man

The MN blood group has three phenotypes M, MN and N with the genotypes L^ML^M, L^ML^N and L^NL^N, respectively. In a sample of 100 individuals, the following numbers with the M, MN and N blood groups were obtained. For simplicity, an example with only two alleles and, hence, three possible genotypes of this blood group is considered here.

Phenotype (blood group)	M	MN	N	Total
Genotype	L^ML^M	L^ML^N	L^NL^N	
Number of individuals	60	30	10	100

We can estimate the frequencies of L^M and L^N alleles in the following two ways with slightly different assumptions.

First Method. Each individual is supposed to have two copies of a gene. In homozygotes, two copies of the same allele are present, while heterozygotes have one copy each of the two alleles. Thus each M individual will have two L^M alleles, while each MN individual will have only one L^M allele. Therefore,

The number of L^M alleles in the population

$$= (2 \times \text{number of } M \text{ individuals}) + \text{number of } N \text{ individuals}$$
$$= (2 \times 60) + 30$$
$$= 150$$

and, the proportion of L^M allele in the population

$$= \frac{\text{Number of } L^M \text{ alleles}}{2 \times \text{total number of individuals in the sample}}$$
$$= \frac{150}{2 \times 100} = 0.75$$

Similarly, the number of L^N alleles

$$= (2 \times 10) + 30 = 50$$

and, the proportion of L^N allele

$$= 50/200 = 0.25$$

Therefore, the frequencies of L^M and L^N alleles are 0.75 and 0.25, respectively.

Second Method. It is assumed that each individual of the sample (or population) produces equal number (say, one) of gametes. In case of the two homozygotes (L^ML^M and L^NL^N), each individual will produce only one type of gamete, while one half of the gametes from each heterozygote (L^ML^N) will have one allele (say, L^M) and the remaining half will contain the other allele (say, L^N). Therefore,

Number of gametes having allele L^M = Number of M Individuals + 1/2 (number of MN individuals)

(Remember, each individual produces only one gamete)

$$= 60 + 1/2\,(30) = 75$$

and, the proportion of L^M gametes

$$= \frac{\text{Number of } L^M \text{ gametes}}{\text{Total number of gametes (= number of individuals in the sample)}}$$
$$= 75/100 = 0.75$$

Similarly, the number of L^M gametes = $10 + 1/2\,(30) = 25$

and, the proportion of L^N gametes = $25/100 = 0.25$

It may be noted that the frequencies of L^M *and* L^N *alleles estimated from the two methods are, as a rule, always the same.*

If a gene has two alleles, *e.g.*, L^M and L^N, their frequencies may be represented by p and q for a generalized treatment. In Table 40.1, $p = 0.75$, $q = 0.25$, and $p + q = (0.75 + 0.25) = 1$. This relationship between p and q, *viz.*, $p + q = 1$, will always hold good due to their definition (that is, p and q are the proportions of the two alleles of a gene in a sample from a population). Therefore, $1 - q = p$, $1 - p = q$ and $1 - (p + q) = 0$. Thus if the frequency of one allele of a gene is known, the frequency of the other allele is readily estimable ($= 1 - p$ or $1 - q$).

40.3. GENOTYPE FREQUENCY

The proportions of different genotypes for a gene is a population are known as ***genotypic frequencies*** for that gene; often they are also called ***zygotic frequencies.*** Genotypic frequencies, like gene frequencies are estimated form samples drawn from random mating populations. The proportion of a genotype in a sample will be the ratio of the number of individuals having that genotype to the total number of individuals in the sample. Thus the genotypic frequencies for MN blood group (Table 40.1) will be as follows: L^ML^M, 0.60 (= 60/100); L^ML^N, 0.30 (= 30/100); and L^NL^N, 0.10 (= 10/100). Gene frequencies are readily estimated from genotypic frequencies using them in place of the numbers of individuals in different genotypic classes. If we follow method 2 of Table 40.1, the frequency of L^M allele will be 0.75 [$= 0.60 + (1/2 \times 0.30)$], while that of L^N allele will be 0.25 [$= 0.10 + (1/2 \times 0.30)$]; these are the same as those estimated earlier.

Gene Pool

- Sum total of alleles of all the genes present in reproductive individuals of a population.
- May be viewed as all the gametes produced by a population (*gametic pool*).
- Next generation arises from a random sample of gametes drawn from the gametic pool.
- Basic feature of Mendelian populations.
- *Features*:
  1. Little affected by contribution, or lack of it, of a few individuals.
  2. Tends to resist changes due to evolutionary factors.

40.4. GENE POOL

The sum total of genes present in the reproductive individuals (or in the gametes) of a population constitutes its ***gene pool,*** while the sum total of gametes produced by the population is called its ***gametic pool.*** In fact, gene pool and gametic pool are almost synonymous since genes are transmitted from one generation to the next through gametes only. *It is assumed that, unless otherwise stated, all the individuals of a Mendelian population contribute equally to its gene and gametic pools.* The concept of gene and gametic pools is based on the fact that in such populations each male gamete is equally likely to unite with any female gamete produced by the population. It may, therefore, be visualised that all the male and female gametes of a population form a pool of gametes. Random samples of male and female gametes from this pool unite randomly to produce the zygotes/individuals of the next generation.

It becomes obvious from the concept of gene pool that if one or few individuals do not contribute to the gametic pool of a population, there will not be a drastic change in its genetic composition. Therefore, the gene and genotype frequencies of the next generation also will not be greatly altered.

Hardy - Weinberg Law

- Independently by Hardy (1908) and Weinberg (1909).
- Gene and genotype frequencies remain constant over generations provided there is no
  1. Selection,
  2. Migration,
  3. Mutation, or
  4. Random drift.
- Frequencies of genotypes *AA*, *Aa* and *aa* are p^2, $2pq$ and q^2, respectively. These are equilibrium frequencies.

40.5. HARDY-WEINBERG LAW

The Hardy-Weinberg law is the fundamental law of population genetics and provides the basis for studying Mendelian populations. This law was developed independently by Hardy (1908) in England and Weinberg (1909) in Germany. ***The Hardy-Weinberg law*** *states that gene and genotype frequencies in a Mendelian population remain constant generation after generation if there is no selection, mutation, migration or random drift. Further, frequencies of the three genotypes for a locus with two alleles , say,* A *and* a, *will remain constant at* p^2 AA, 2pq Aa *and* q^2 aa, *where* p *and* q *are the frequencies of alleles* A *and* a, *respectively.* The validity of Hardy-Weinberg law can be demonstrated by assuming either random union of gametes or random mating among genotypes.

40.5.1. Ramdom Union of Gametes

Let us suppose that the two alleles of a gene are A and a, their frequencies are p and q, respectively, and that A and a (male and female) gametes unite at random to yield zygotes. Since the union of male and female gametes is random, *the frequency of union of a male gamete with a female gamete will be the product of their individual frequencies in the population.* Thus the frequency of union between A male and A female gametes will be p^2 ($= p \times q$), between a male and A female gametes will be pq ($= p \times q$), between A male and a female gametes will also be pq ($= q \times p$), and between a male and a female gametes will be q^2 ($= q \times q$). These unions will yield the genotypes *AA*, *Aa* and *aa;* the frequencies of these genotypes will be p^2, $2pq$ and q^2, respectively (Table 40.2).

The three genotypes, *AA*, *Aa*, and *aa*, will produce A and/or a gametes; the frequencies of these gametes can be estimated as follows. Each individual of the population is assumed to produce equal number of gametes, say, 1. It can be seen that,

$$p^2 + 2pq + q^2 = (p + q)^2$$

Since, $p + q = 1$

Hence, $p^2 + 2pq + q^2 = 1$

TABLE 40.2

Gene and genotype frequencies for a gene obtained from a random union of gametes in a *Mendelian population*

Suppose,	Genotype of gamete		*A*	*a*
	Frequency of gamete		*p*	*q*
	Random union among the gametes will yield,			
		Male gamete	*A*	*a*
	Female gamete		(*p*)	(*q*)
	A (*p*)		*AA*	*Aa*
			(p^2)	(*pq*)
	a (*q*)		*Aa*	*aa*
			(*pq*)	(q^2)
	Thus, Genotype of zygote	*AA*	*Aa*	*aa*
	Frequency of zygote	p^2	2*pq*	q^2

Note : The frequencies of union between two gametes will be equal to the product of their individual frequencies.

Thus the size of population is also 1. All *AA* individuals will produce only *A* gametes, *Aa* individuals will produce 1/2 *A* and 1/2 *a* gametes, and *aa* individuals will give rise to only *a* gametes. Therefore, the frequency of *A* gametes will be,

= All the gametes from *AA* individuals + 1/2 the gametes from *aa* individuals
= $p^2 + 1/2\ (2pq)$
= $p^2 + pq$
= $p\ (p + q)$ (*p* is taken common)
= p (since, $p + q = 1$)

Similarly, the frequency of *a* gametes produced by the population will be,

= All the gametes from *aa* + 1/2 the gametes from *Aa* individuals
= $q^2 + 1/2\ (2pq)$
= $q^2 + pq$
= $q\ (p + q)$ (*q* is taken common)
= q (since, $p + q = 1$)

Obviously, the frequencies of *A* and *a* gametes produced by the new generation of the population are the same (*p* and *q*) as those of the parent population. A random union among these gametes will again yield the three genotypes *AA*, *Aa* and *aa* in the frequencies p^2, 2*pq* and q^2, respectively (Table 40.2); these frequencies will be the same as those of the present population.

40.5.2. Random Mating among Genotypes

Let us suppose that a Mendelian population has the genotypes *AA*, *Aa* and *aa* in the frequency p^2, 2*pq* and q^2, respectively. Since the matings among genotypes are assumed to be random, *the frequency of mating between males and females of any given genotype(s) will be the product of the frequencies of the concerned genotypes in the population.* Therefore, the frequencies of matings in which both males and

females have the same genotype, *e.g.*, ♀ *AA* × ♂ *AA*, *Aa* × *Aa* and *aa* × *aa*, will be equal to the square of the frequencies of the concerned genotypes (*e.g.*, $p^2 \times p^2 = p^4$, $2pq \times 2pq = 4p^2q^2$, and $q^2 \times q^2 = q^4$, respectively). But the frequencies of those matings in which males and females have different genotypes, *e.g.*, *AA* × *Aa*, *AA* × *aa*, and *Aa* × *aa*, will be twice the product of the frequencies of the two genotypes involved in the mating (*e.g.*, $2 \times p^2 \times 2pq = 4p^3q$, $2 \times p^2 \times q^2 = 2p^2q^2$, and $2 \times 2pq \times q^2 = 4pq^3$, respectively). This is because such matings can occur in two different ways: (1) females of one genotype and males of the other, and (2) *vice-versa*. For example, the two ways for the mating *AA* × *Aa* are; (1) ♀ *AA* × ♂ *Aa* and (2) ♀ *Aa* × ♂ *AA*. Therefore, the frequency of this mating will be the sum of the frequencies of these two mating combinations [= ($p^2 \times 2pq$) + ($2pq \times p^2$) = $2(p^2 \times 2pq) = 4p^3q$)].

The frequencies of different matings among the three genotypes, and those of their progeny are summarised in Table 40.3. The frequencies of different progenies from a mating are deduced by simple logic. The frequency of progeny from a mating will be proportionate to the frequency of that mating; it is assumed here that the rate of reproduction in all the matings is equal. Therefore, the proportions of progeny having different genotypes from a mating can be obtained by partitioning the frequency of that mating in accordance with the proportions of the progeny of different genotypes expected from the mating. For example, mating *AA* × *AA* would produce only *AA* progeny. Hence this mating will produce *AA* progeny with a frequency of p^4 as this is the frequency of this mating. But the mating *Aa* × *aa* would yield two types of progeny, *viz.*, *Aa* and *aa*, in equal frequency. Since the frequency of this mating is $4p^3q$, the frequencies of the two types of progenies will be $2p^3q$ each, *i.e.*, $2p^3q$ *Aa* and $2p^3q$ *aa*. In this manner, the frequencies of different genotypes in the progeny can be readily estimated (Table 40.3).

TABLE 40.3

Gene and genotype frequencies for a gene obtained from random mating among different genotypes.

Suppose,	Genotype	*AA*	*Aa*	*aa*
	Frequency	p^2	2*pq*	q^2

Random mating among the above three genotypes would yield,

Mating	*Probability*	*Frequency of progeny*		
		AA	Aa	aa
AA × *AA*	p^4	p^4		
AA × *Aa*	$4p^3q$	$2p^3q$	$2p^3q$	
AA × *aa*	$2p^2q^2$		$2p^2q^2$	
Aa × *Aa*	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
Aa × *aa*	$4pq^3$		$2pq^3$	$2pq^3$
aa × *aa*	q^4			q^4

Note : 1. The frequency of a mating in which both males and females have the same genotype is equal to the square of the frequency of the genotype concerned.
2. A mating in which males and females have different genotypes has the frequency, which is twice the product of the frequencies of the genotypes involved in the mating.
3. The frequency of different progeny obtained from a mating is deduced by partitioning the frequency of the mating in accordance with the proportions of the different genotypes expected from that mating.

The frequencies of different genotypes in the progeny obtained by random mating among the genotypes can now to determined. The frequency of *AA* progeny would be,

$= p^4 + 2p^3q + p^2q^2$
$= p^2 (p^2 + 2pq + q^2)$ (p^2 is taken common)
$= p^2$ (since, $p^2 + 2pq + q^2 = 1$)

Similarly, the frequency of *aa* progeny will be,

$= p^2q^2 + 2pq^3 + q^4$
$= q^2 (p^2 + 2pq + q^2)$ (q^2 is taken common)
$= q^2$ (since, $p^2 + 2pq + q^2 = 1$)

And the frequency of *Aa* progeny will be,

$= 2p^3q + 2p^2q^2 + 2p^2q^2 + 2pq^3$
$= 2p^3q + 4p^2q^2 + 2pq^3$
$= 2pq (p^2 + 2pq + q^2)$ ($2pq$ is taken common)
$= 2pq$ (since, $p^2 + 2pq + q^2 = 1$)

Obviously, random mating among the genotypes *AA*, *Aa* and *aa* having the frequencies of p^2, $2pq$ and q^2, respectively, yields the same three genotypes in the same frequencies, (p^2 *AA*, $2pq$ *Aa* and q^2 *aa*).

Restoration of Hardy - Weinberg Equilibrium

- *One gene, 2 alleles*: One generation.
- *One gene, many alleles*: One generation.
- *Two genes*: Several generations; restoration gradual.
- *Linked genes*: Restoration slower; depends on recombination frequency.
- *Sex-linked genes*: Restoration slow; may take (6-7) generations.

40.6. HARDY-WEINBERG EQUILIBRIUM

According to the Hardy-Weinberg law, gene and genotype frequencies in Mendelian populations remain constant generation after generation if there is no migration, mutation, selection or random drift. In such populations, the three genotypes for a gene (say, *A/a*), *e.g.*, *AA*, *Aa* and *aa*, are present in the frequencies of p^2, $2pq$ and q^2, respectively. Such populations are said to be at an ***equilibrium*** or at ***Hardy-Weinberg equilibrium*** since their gene and genotype frequencies are constant from one generation to the next. Further, in these populations the three genotypes for a gene, *e.g.*, *AA*, *Aa* and *aa*, are present in the frequency p^2, $2pq$ and q^2, respectively; these frequencies denote ***Hardy-Weinberg equilibrium frequencies.*** *The existence of Hardy-Weinberg equilibrium in a population shows the absence of migration, mutation, selection and random drift in them.*

40.6.1. Equilibrium for One Gene with Two Alleles

Whether a population is at equilibrium for a gene with two alleles is determined by (1) studying a random sample drawn from it. (2) The individuals in the sample are classified into the three genotypes possible for the gene. (3) From this data frequencies of the two alleles of the gene are estimated. (4) These gene frequencies

are then used to compute the equilibrium frequencies of the three genotypes for the gene according to the general expectation of p^2 *AA*, $2pq$ *Aa* and q^2 *aa*. (5) Based on these estimated genotype frequencies, the expected numbers of individuals with the different genotypes are estimated; the numbers will be $p^2 \times N$, *AA*; $2pq \times N$, *Aa*; and $q^2 \times N$, *aa*; where N is the total number of individuals in the sample. (6) Now a χ^2-test is carried out to see if the observed frequencies of the three genotypes are in agreement with the expected ones (Table 40.4). *If the calculated value of* χ^2 *is nonsignificant, the population is at equilibrium for this gene, but if the* χ^2 *value is significant the population is not at equilibrium.*

The equilibrium may be disturbed by migration, mutation etc., in which case the frequencies of the three genotypes will deviate from p^2 AA, 2pq Aa *and* q^2 aa. *In the case of a single gene with two alleles, equilibrium is restored (i.e., genotypic frequencies become* p^2 AA, 2pq Aa *and* q^2 aa*) in one generation after the disturbing factor has been removed.*

40.6.2. Equilibrium for One Gene with Multiple Alleles

Many genes are known to have multiple alleles. Let us assume that a gene *A* has three alleles A_1, A_2 and A_3. There are six possible genotype for this gene, *viz.*, A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3, and A_3A_3. The frequencies of alleles A_1, A_2, and A_3 can be estimated from the frequencies of these six genotypes present in a random sample drawn from the population, the principle remaining the same as that in the case of two alleles. Let us suppose that the frequencies of alleles A_1, A_2 and A_3 are p, q and r, respectively, so that $p + q + r = 1$. Since each gamete will contain only one of these three alleles, p, q and r will also be the frequencies of A_1, A_2 and A_3 gametes, respectively.

TABLE 40.4

Chi-square test of equilibrium for L^M and L^N alleles (for MN blood groups) in a sample from Beduins

Blood group	M	MN	N	Total
Genotype	L^ML^M	L^ML^N	L^NL^N	
Observed number	119	76	13	208

Frequency of L^M allele in the sample (p) = (M + 1/2 MN) / Total number
= (119 + 1/2 × 76)/208
Hence, p = 157/208 = 0.75 (approx.)
And, frequency of L^N allele (q) = (N + 1/2 MN) / Total number
= (13 + 1/2 × 76)/208
Hence, q = 51/208 = 0.25 (approx.)

At equilibrium, the frequencies of M, MN and N blood groups will be p^2, $2pq$ and q^2, respectively. Therefore, equilibrium frequency of

M blood group = $p^2 = 0.75^2 = 0.56$
MN blood group = $2pq = 2(0.75)(0.25) = 0.38$
and, N blood group = $q^2 = 0.25^2 = 0.06$

The expected numbers of individuals with M, MN and N blood groups in the sample, if the population were at equilibrium, will be p^2, $2pq$ and q^2, respectively, each multiplied by the number of individuals in the sample. Therefore, the expected number of individuals in the sample having,

M blood group = $0.56 \times 208 = 116.5$
MN blood group = $0.38 \times 208 = 79.0$
and, N blood group = $0.06 \times 208 = 12.5$

Now, a χ^2-test is done to test if the observed frequencies of M, MN and N blood groups agree with their expected frequencies.

Blood group	M	MN	N	Total
Observed frequency (O)	119	76	13	208
Expected frequency (E)	116.5	79	12.5	208
Deviation (O–E)	+2.5	–3.0	+0.5	0
$(O-E)^2/E$	0.054	0.114	0.020	0.188

$\chi^2 = 0.118$

This χ^2 value has 1 degrees of freedom because the expected numbers of the three genotypes are based on (1) the frequencies of L^M and L^N alleles, and (2) the total number of individuals in the sample. This permits the selection of an arbitrary number for only one of the three genotypes; the frequencies of the other two genotypes have to be adjusted to obtain the given frequencies with the same number of observations. For example, if we select 150 as the number of M individuals in the above example, the frequencies of MN and N blood groups have to be 12 and 46, respectively; if we alter the frequencies of MN and N individuals in any way, the gene frequencies and/or the total number of individuals in the population will have to be changed (which must not happen).

Here, *the calculated value of χ^2 is less than the table value at 0.05 probability and 1 d.f., indicating that the population is at equilibrium for this gene.*

A random union among A_1, A_2 and A_3 male and female gametes will yield the following zygotes or genotypes: $p^2\ A_1A_1$, $2pq\ A_1A_2$, $2pr\ A_1A_3$, $q^2\ A_2A_2$, $2qr\ A_2A_3$ and $r^2\ A_3A_3$. This can be verified by making a checker-board with $p\ A_1$, $q\ A_2$ and $r\ A_3$ gametes both on the male and female sides and scoring the frequencies of the different genotypes obtained from union among them in all possible combinations. The same frequencies are obtained by expansion of the trinomial $(p + q + r)^2$. These frequencies are the equilibrium frequencies of the six genotypes for the gene *A*. If the number of alleles is more, the appropriate expansion contains a corresponding number of terms. A chi-square test is now carried out to determine if the observed frequencies agree with the expected ones. In this case also, equilibrium is restored in one generation of the removal of the disturbing factor.

40.6.3. Equilibrium for Two Genes

When two genes are considered together, equilibrium is not restored in one generation, but the genotypic frequencies approach equilibrium slowly in several generations. Let us assume that alleles *A* and *a* of one gene have the frequencies *p* and *q*, respectively, while alleles *B* and *b* of another gene have the frequencies *r* and *s*, respectively. Independent segregation of these two genes (*A/a* and *B/b*) will yield the following gametes: *AB, Ab, aB* and *ab*; at equilibrium, the frequencies of these gametes will be *pr, ps, qr* and *qs*, respectively. A random union among these gametes will yield 9 different genotypes; the frequencies of these genotypes can be obtained by expanding the quadrinomial $(pr + ps + qr + qs)^2$. The nine genotypes and their equilibrium frequencies will be as follows: p^2r^2 *AA BB*, $2p^2rs$ *AABb*, $2pqr^2$ *AaBB*, $4pqrs$ *AaBb*, p^2s^2 *AA bb*, $2pqs^2$ *Aa bb*, q^2r^2 *aa BB*, $2q^2rs$ *aa Bb* and q^2s^2 *aa bb*.

These frequencies can also be obtained by making a checker-board and writing p^2 *AA*, $2pq$ *Aa* and q^2 *aa* on one side and r^2 *BB*, $2rs$ *Bb* and s^2 *bb* on the other; a random union of the three genotypes for gene *A* and those for gene *B* will yield the same result (Table 40.5).

TABLE 40.5

The equilibrium frequencies of different genotypes produced by independent segregation of two genes *A* and *B*. The gene frequencies are, *p A, q a, r B* and *s b*, so that $p + q = 1$ and $r + s = 1$.

Genotypes and frequencies	*AA* p^2	*Aa* $2pq$	*aa* q^2
BB	*AA BB*	*Aa BB*	*aa BB*
r^2	p^2r^2	$2pqr^2$	q^2r^2
Bb	*AA Bb*	*Aa Bb*	*aa Bb*
$2rs$	$2p^2rs$	$4pqrs$	$2q^2rs$
bb	*AA bb*	*Aa bb*	*aa bb*
s^2	p^2s^2	$2pqs^2$	q^2s^2

Gene frequencies estimated from a random sample are used for estimating the equilibrium frequencies of the nine genotypes for the two genes. A chi-square test is then carried out to see if the observed frequencies of these genotypes in the sample are in agreement with their expected frequencies.

Two gametes, *viz., AB* and *ab*, out of the four possible gametes produced by the segregation of *A/a* and *B/b* (*AB, Ab, aB* and *ab*) are in coupling phase, while the remaining two, *viz., aB*, and *Ab*, are in the repulsion phase. *(Here, the terms coupling phase and repulsion phase do not reflect linkage between genes* A *and* B*; they only reflect the types of allelic combinations.)* At equilibrium, the product of frequencies of the coupling phase gametes is equal to that of the frequencies of repulsion phase ones. That is,

$$fAB \times fab = fAb \times faB$$

where, *fAB*, etc. represent the frequencies of *AB*, etc. gametes. If the population is not at equilibrium for the two genes, $fAB \times fab$ will not be equal to $fAb \times faB$; the difference between these two values is called ***disequilibrium,*** and is represented by *d*. Thus,

$$d = (fAb \times faB) - (fAB \times fab)$$

If the value of *d* is positive, the frequency of each coupling phase gamete (*AB* and *ab*) will increase by *d* at equilibrium, while that of each repulsion phase gamete (*Ab* and *aB*) will decrease by the same value. The reverse will be the case if *d* is negative. After each generation of random mating following the removal of disturbing factor(s), the value of *d* becomes 1/2 of that in the previous generation. This is because, half of the gametes produced by independent segregation for two genes are coupling and the remaining half are repulsion phase gametes. Thus after four generations, 93.75% of the *d* will be dissipated, and after six generations only 1.6% of the *d* shall remain (Table 40.6).

TABLE 40.6

The rate of dissipation of *d* (*disequilibrium*) in the frequencies of the four gametes produced by independent segregation of two genes (*A/a* and *B/b*)

Gamete	Change in frequency after generation						
	1	*2*	*3*	*4*	*5*	*6*	*Equilibrium*
AB	+ 0.5 *d*	+ 0.75 *d*	+ 0.875 *d*	+ 0.938 *d*	+ 0.969 *d*	+ 0.984 *d*	+ *d*
Ab	− 0.5 *d*	− 0.75 *d*	− 0.875 *d*	− 0.938 *d*	− 0.969 *d*	− 0.984 *d*	− *d*
aB	− 0.5 *d*	− 0.75 *d*	− 0.875 *d*	− 0.938 *d*	− 0.969 *d*	− 0.984 *d*	− *d*
ab	+ 0.5 *d*	+ 0.75 *d*	+ 0.875 *d*	+ 0.938 *d*	+ 0.969 *d*	+ 0.984 *d*	+ *d*

Note: In any generation, the value of *d* will be one-half of its value in the previous generation.

40.6.4. Effect of Linkage on Equilibrium

When two genes are linked, the difference between the products of frequencies of the two repulsion phase gametes and the two coupling phase gametes is called ***linkage disequilibrium*** and is designated by *d*. In each generation, the value of *d* will decline by the product of the frequency of recombination between the two genes and the value of *d* in the previous generation. Since the recombination frequency will almost always be less than 0.5 (which is the value in the case of independent segregation), the rate of dissipation of linkage disequilibrium will always be much lower than that in the case of independent segregation; this rate shall depend solely on the frequency of recombination (Table 40.7). However, at equilibrium, the frequencies of the different gametes shall change by *d*, which is the same as in the case of independent segregation.

40.6.5. Equilibrium for Sex-Linked Genes

Sex-linked genes are located in the *X* chromosomes; human males (*XY*) have one *X*, while females (*XX*) have two *X* chromosomes. Therefore, a sex-linked gene (say *A*) will have the genotypes *A* and *a* in males, while females will have the genotypes *AA*, *Aa* and *aa*. If the population is at equilibrium for this gene, and the frequencies of alleles *A* and *a* are *p* and *q*, respectively, the frequencies of genotypes *A* and *a* in the males will be *p* and *q*, respectively. Similarly, the frequencies of genotypes *AA*, *Aa* and *aa* in the females will be p^2, $2pq$ and q^2, respectively. In such a population, the gene and genotype frequencies shall remain constant generation after generation since in both males and females the frequencies of *A* and *a* gametes will be *p* and *q*, respectively.

It may be noted that male zygotes receive their *X* chromosome from females, while female zygotes receive one *X* from males and one *X* from females. Therefore, males of a generation exhibit the gene frequencies of females of the previous generation. But the gene frequencies in females of a generation are the average of those in the males and the females of the previous generation. Let us suppose that a population is not at equilibrium for a gene (say, *A/a*), and that frequencies of its two alleles (say, *A* and *a*) are p_m and q_m, respectively, in males, and p_f and q_f, respectively, in females of a generation. In the next generation, males will exhibit alleles *A* and *a* in frequencies p_f and q_f, respectively. But the frequencies of these alleles will be $(p_f + p_m)/2$ for *A* and $(q_f + q_m)/2$ for *a* in the case of females. Therefore, a population approaches equilibrium for a single sex-linked gene rather slowly; it takes about 6-7 generations for this purpose. At equilibrium, both males

TABLE 40.7

The effect of recombination frequency between two genes on the rate of change in *d* in a random mating population

Generation	Recombination frequency				
	0.5 (Independent segregation)	*0.4*	*0.3*	*0.2*	*0.1*
	Change in *d*				
1	0.5 *d*	0.4 *d*	0.3 *d*	0.2 *d*	0.1 *d*
2	0.75 *d*	0.64 *d*	0.51 *d*	0.36 *d*	0.19 *d*
3	0.875 *d*	0.784 *d*	0.657 *d*	0.488 *d*	0.271 *d*
4	0.938 *d*	0.870 *d*	0.760 *d*	0.590 *d*	0.344 *d*
5	0.969 *d*	0.922 *d*	0.832 *d*	0.672 *d*	0.410 *d*
6	0.984 *d*	0.953 *d*	0.882 *d*	0.738 *d*	0.459 *d*
Equilibrium	*d*	*d*	*d*	*d*	*d*

Note : In each generation, the value of *d* will decline by the product of the recombination frequency and the value of *d* in the previous generation.

and females of the population show the average of gene frequencies in the two sexes, which are estimated as follows.

$$\hat{p} = (p_m + 2p_f)/3$$

$$\text{and, } \hat{q} = (q_m + 2q_f)/3$$

where, $\hat{p}$ and $\hat{q}$ are equilibrium or average frequencies of the alleles *A* and *a*, respectively. Here, $2p_f$ and $2q_f$ are used since each female has two *X* chromosomes, while each male has only one *X* chromosome. Therefore, the total number of *A* alleles in the population will be $p_m + 2p_f$, and that of *a* allele will be $q_m + 2q_f$.

40.6.6. Equilibrium for Genes Showing Dominance

In case of completely dominant genes, the homozygotes for the dominant allele and the heterozygotes have the same phenotype. Therefore, such genes produce only two phenotypes: (1) dominant and (2) recessive phenotypes. For example, a completely dominant gene *T* determines the ability of human beings to detect the bitter taste of dilute solutions of phenylthiocarbamide (PTC; or phenylthiourea). Individuals with genotypes *TT* or *Tt* are able to taste dilute solutions of PTC (tasters), while those with the genotype *tt* are unable to do so (nontasters). Obviously, we cannot determine the frequencies of *TT* and *Tt* genotypes as both produce the same 'taster' phenotype. But the frequency of *tt* genotype will be the same as that of the nontaster phenotype; at equilibrium this will be q^2 from which the value of q $(= \sqrt{q^2})$ and p $(= 1 - q)$ can be easily estimated. Thus if the frequency of nontasters in a sample were 0.16, $q^2 = 0.16$, $q = \sqrt{0.16} = 0.4$, and $p = 1 - q = 1 - 0.4 = 0.6$.

The test for equilibrium for such genes is based on two ***population ratios*** proposed by Synder. These population ratios can be estimated if the following data were available.

1. The numbers of matings involving the different phenotypic combinations.
2. The numbers of progeny exhibiting different phenotypes produced from these matings.

At equilibrium, the genotypes *TT*, *Tt* and *tt* will have the frequency p^2, $2pq$ and q^2, respectively. The different types of possible matings, their frequencies, and the phenotypes of the parents are summarised in Table 40.8. Clearly, taster × taster matings having the frequency of $p^4 + 4p^3q + 4p^2q^2$, will yield nontaster progeny only when both the parents are heterozygous, *i.e.*, $Tt \times Tt$; such matings have the frequency of $4p^2q^2$, and 1/4 of their progeny (= p^2q^2) will be nontasters (*tt*). Thus, out of the total of $p^4 + 4p^3q + 4p^2q^2$ progeny produced by taster × taster matings, only p^2q^2 progeny will be nontasters (*i.e.*, will have the recessive phenotype). The ratio between these two values (*i.e.*, the number of nontaster progeny from taster × taster matings, and the total number of progeny produced by such matings) is the ***first population ratio*** of Synder; on simplification, it will become $q^2/(1 + q)^2$ as shown below.

First population ratio = $p^2q^2/(p^4 + 4p^3q + 4p^2q^2)$
= $p^2q^2/p^2(p^2 + 4pq + 4q^2)$ (p^2 is taken common)
= $q^2/(p^2 + 4pq + 4q^2)$ (both divided by p^2)
= $q^2/[(1 - q)^2 + 4(1 - q)(q) + 4q^2]$ (on replacing p with $1 - q$).
= $q^2/(1 - 2q + q^2 + 4q - 4q^2 + 4q^2)$
= $q^2/(1 + 2q + q^2)$
= $q^2/(1 + q)^2$ [since $(1 + q)^2 = 1 + 2q + q^2$)]

TABLE 40.8

The equilibrium frequencies of different types of matings among the genotypes and phenotypes for a single dominant gene *T* (*T* determines taster of *PTC*, and *t* governs nontaster of *PTC*; $fT = p$, and $ft = q$)

Female\male		*Taster (dominant)*		*Non-taster (recessive)*
		TT	Tt	tt
		p^2	2pq	q^2
Taster	TT p^2	p^4	$2p^3q$	p^2q^2
	Tt 2pq	$2p^3q$	$4p^2q^2$*	$2pq^3$*
Nontaster	tt q^2	p^2q^2	$2pq^3$*	q^4*

* Only these matings will produce progeny with the recessive phenotype. From $Tt \times Tt$ matings (frequency, $4p^2q^2$), 1/4 of the progeny will be *tt* or nontaster (frequency, p^2q^2); from $Tt \times tt$ matings (frequency, $4pq^3$), 1/2 of the progeny will be *tt* (frequency, $2pq^3$); while from $tt \times tt$ matings (frequency, q^4), all the progeny will be *tt* (frequency, q^4).

The above may be summarised as follows:

Mating	*Frequency*	*Frequency of progeny with the recessive phenotype*
Dominant × Dominant	$p^4 + 4p^3q + 4p^2q^2$	1/4 $(4p^2q^2)$* = p^2q^2
Dominant × Recessive	$2p^2q^2 + 4pq^3$	1/2 $(4pq^3)$** = $2pq^3$
Recessive × Recessive	q^4	q^4

*Recessive progeny from only $Tt \times Tt$ matings.
**Recessive progeny from only $Tt \times tt$ matings.

Similarly, taster × nontaster (or dominant × recessive) matings will produce a total of $2p^2q^2 + 4pq^3$ progeny. Of these, $2pq^3$ progeny will be nontasters (that is, 1/2 of the progeny produced by $Tt \times tt$ matings). The ratio of the frequency of recessive progeny from dominant × recessive matings to that of the total number of progeny from such matings is the ***second population ratio*** of Synder. The second population ratio is, therefore, $2pq^3/(2p^2q^2 + 4pq^3)$, which yields $q/(1 + q)$ on simplification.

Second population ratio = $2pq^3/(2p^2q^2 + 4pq^3)$
= $2pq^3/[2pq^2(p + 2q)]$ ($2pq^2$ is taken common)
= $q/(p + 2q)$ (both divided by $2pq^2$)
= $q/(p + q + q)$ ($2q$ written as $q + q$)
= $q/(1 + q)$ (since $p + q = 1$)

TABLE 40.9

Test for equilibrium for a dominant gene *T* (*T*, taster for PTC; *t*, nontaster)

Mating	*Number of pairs*	*Number of parents*	*Number of progeny*			*Population ratio*	
			Taster	*Nontaster*	*Total*	*Observed*	*Expected*
Taster × taster (D × D)*	425	850	929	130	1,059	0.123	0.122
Taster × nontaster (D × R)	289	578	483	278	761	0.365	0.349
Nontaster × nontaster (R × R)	86	172	(5)**	218	223		
Total	800	1,600	1,417	626	2,043		

* D, dominant; R, recessive.
** Possible explanation: illegitemate progeny, erroneous diagnosis or due to some other reason; generally, unexpected.

Number of nontaster parents = 289 + 2(86) = 461
Frequency of nontaster parents (q^2) = 461/1600 = 0.2881
Therefore, $q = \sqrt{q^2} = \sqrt{0.2881} = 0.537$
Estimate of the first population ratio = $q^2/(1 + q)^2$
= $(0.537)^2/(1 + 0.537)^2 = (0.537)^2/(1.537)^2$
= 0.2883/2.3624 = 0.122
Similarly, estimate of the second population ratio
= $q/(1 + q)$
= 0.537/(1 + 0.537) = 0.537/1.537
= 0.349

Note: *These estimates of the two population ratios are the ones that will be obtained if the population were at equilibrium for the gene* T/t.

The observed value of the first population ratio

$$= \frac{\text{Number of recessive progeny from D} \times \text{D matings}}{\text{Total number of progeny from such matings}}$$

= 130/1,059 = 0.123

And, observed value of second population ratio

$$= \frac{\text{Number of recessive progeny from D} \times \text{R matings}}{\text{Total number of progeny from such matings}}$$

= 278/761 = 0.365

Conclusion: The observed and expected values of the two population ratios are comparable; therefore, the population is at equilibrium for this gene.

The procedure for testing equilibrium for a completely dominant gene is as follows.

1. A random sample is drawn from the population and the frequencies of dominant and recessive phenotypes are determined.
2. The value of q is estimated as the square root of the frequency of recessive phenotype.
3. Now the values of the two population ratios of Synder are estimated using this value of q; these are the expected values of the two population ratios if the population were at equilibrium.
4. The observed values for these ratios are estimated as the ratios of the numbers of recessive progeny to the total numbers of progeny produced by *dominant* × *dominant* (*first ratio*) and *dominant* × *recessive* (*second ratio*) matings, respectively, among the individuals of the sample.
5. If the observed and estimated values of the two population ratios are comparable, the population is considered to be at equilibrium for the gene, otherwise not (Table 40.9).

QUESTIONS

1. Explain the Hardy-Weinberg law assuming random union of gametes (or random mating among genotypes).
2. Define Hardy-Weinberg equilibrium. Describe the procedure for determining if a Mendelian population is at equilibrium for a gene with two alleles.
3. Explain the effects of following on Hardy-Weinberg equilibrium in a Mendelian population: (*i*) multiple alleles at a locus, (*ii*) two genes, (*iii*) linkage and, (*iv*) sex-linkage.
4. Define gene frequency. Briefly describe the procedure for estimation of gene frequency in a given population.
5. Describe the procedure for testing if a Mendelian population is at equilibrium for a gene at which one of the two alleles shows complete dominance.
6. Write short notes on the following:
   (*i*) Mendelian population, (*ii*) gene and genotype frequencies,
   (*iii*) gene pool, (*iv*) Hardy-Weinberg law,
   (*v*) Hardy-Weinberg equilibrium, (*vi*) population ratios of Snyder.

41

Population Genetics
II. Factors Affecting Gene Frequencies

41.1. INTRODUCTION

If gene and genotype frequencies remained constant generation after generation (following the Hardy-Weinberg law) in natural populations, the populations would have constant genetic composition and there would be no evolution in them. It is generally accepted that most natural populations are constantly evolving. The gene and genotype frequencies in natural populations, therefore, must undergo change to permit evolution. Gene frequencies in Mendelian populations change due to one or more of the following factors: (1) migration, (2) mutation, (3) selection and (4) random drift; these factors are, therefore, called ***evolutionary factors*** or ***evolutionary forces*** because they lead to evolution in natural populations.

Evolutionary Factors

- Gene and genotype frequencies change due to:
  1. migration,
  2. mutation,
  3. selection, and
  4. random drift.
- Change in gene frequencies is essential for evolution.

41.2. MIGRATION

Migration refers to the movement of some individuals of a population into another population and their intermating with the individuals of the latter. *Thus migration is the contribution to the gene pool of a population by some individuals of another population.* Migration may lead to a change in gene frequency depending upon the following.

1. The proportion of migrant individuals to the total number (after migration) of individuals in the population; this is called ***migration ratio.***
2. The difference between the gene frequencies of the migrant individuals and that of the population into which migration has occurred.

The migration ratio (m) is estimated as follows.

$$m = M/(N + M)$$

where, M is the number of migrant individuals, and N is the number of individuals in the population before migration.

If the frequency of allele a in the population before migration is q_o, that in the migrant individuals is q_m, and the migration ratio is m, the frequency of a after migration, *i.e.*, q_1, will be the weighted average of the two frequencies, q_o and q_m.

$$\begin{aligned} q_1 &= q_o(1-m) + mq_m \\ &= q_o - mq_o + mq_m \\ &= q_o + m(q_m - q_0) \qquad (m \text{ is taken common}) \end{aligned}$$

Thus, the amount of change in the value of q, represented by Δq, due to migration in one generation will be,

$$\begin{aligned} \Delta q &= q_1 - q_o \\ &= [q_o + m(q_m - q_o) - q_o \qquad (\text{on replacing } q_1 \text{ by its estimate}) \\ &= m(q_m - q_o) \qquad (\text{since } +q_o \text{ and } -q_o \text{ cancel out}) \end{aligned}$$

But if the migration from a given population were recurrent, *i.e.*, occuring in each generation, the rate of migration were the same in each generation, *i.e.*, the migration ratio, m, had the same value, and migration occurred for n generations, the following equation will be obtained.

$$q_n - q_m = (1-m)^n (q_o - q_m)$$

or,

$$(1-m)^n = (q_n - q_m)/(q_o - q_m)$$

where, q_n is the value of q after n generations of constant recurrent migration from the same population, and n is the number of generations for which migration has taken place. The above relationship is among the following five factors, m, n, q_o, q_m and q_n; if the values of any four of these were known, the value of the fifth factor can be readily estimated.

41.3. MUTATION

A sudden, heritable change (not produced by segregation or recombination) in the characteristics of an organism is called ***mutation;*** they are the ultimate source of all genetic variation. Mutation may be (1) nonrecurrent or (2) recurrent.

41.3.1. Nonrecurrent Mutation

A nonrecurrent mutation takes place only once, that is, the same mutation will not occur in a future generation of a population. For example, one A allele of a AA homozygote may mutate to a yielding the heterozygote Aa. The probability of transmission of the mutant allele a to the next generation will depend on the number of progeny produced by the heterozygote Aa. If Aa does not produce any progeny, a will not be transmitted to the next generation. But if it produces one progeny, the probability of transmission of a will be 1/2, while it will be 3/4 if two progeny were produced. Thus, the probability of transmission of a to the next generation will increase with an increase in the number of progeny produced by the Aa individual.

However, the number of progeny produced by the different individuals of a population is never equal even in the absence of selection; actually, it follows a Poisson distribution. Therefore, the probability of transmission of a has to be corrected according to the Poisson distribution. After this correction, if the average number of progeny produced by a mating in the population is two, the probability of transmission of a by the Aa individual to the next generation will be 0.6321 in place of 0.75; it will be 0.4685 after three generations and only 0.05 after 30 generations.

Thus, *if there is a nonrecurrent mutation in a gene, the mutant allele will be lost in one of the subsequent generations.*

41.3.2. Recurrent Mutation

When a given gene mutation occurs regularly, although, at a very low frequency, it is called ***recurrent mutation.*** If gene A is mutating into a at the rate u in each generation, u proportion of A alleles will change into the a allele. Thus,

The frequency of newly produced a alleles $= p_o u$
The total frequency of a allele after mutation $= q_o + p_o u$
And, the frequency of A allele after mutation $= p_o - p_o u$

where, p_o and q_o are the frequencies of A and a alleles, respectively, in the generation before mutation. Thus in each generation, the frequency of A allele will decline by pu and that of a will increase by the same value, where p is the frequency of A allele in the generation before mutation. Thus the value of q will increase by pu in each generation, *i.e.*, $\Delta q = pu$. It should be noted that the value of p will go on decreasing with every generation. Ultimately, all the A alleles will be converted into a, *i.e.*, a will become 'fixed' in the population. When an allele has the frequency of 1, it is said to have been ***fixed***, since its frequency can no longer change, except through migration and/or mutation.

It most cases, both forward ($A \rightarrow a$) and reverse ($a \rightarrow A$) mutations take place; generally, the rates of forward mutations are much higher than those of reverse mutations. Let us suppose that the rate of mutation of A to a is u and that of a to A is v, and the frequency of A in a generation is p_o and that of a is q_o. Mutation would convert $p_o u$ A alleles into a and $q_0 v$ a alleles into A. Thus q_o will decline by $q_o v$ and gain by $p_o u$. Thus, the value of p after mutation (p_1) will be,

$$p_1 = p_o - p_o u + q_o v$$

And,

$$\begin{aligned} \Delta p &= p_1 - p_o \\ &= (p_o - p_o u + q_o u) - p_o \qquad (\text{since } p_1 = p_o - p_o u + q_o v) \\ &= -p_o u + q_o v = q_o v - p_o u \end{aligned}$$

Thus, p and q will change due to mutation only when $p_o u$ is not equal to $q_o v$. But when $q_o v = p_o u$, the values of p and q will not be affected by mutation, *i.e.*, they will be at equilibrium; such an equilibrium is called ***mutational equilibrium.*** At mutational equilibrium,

$$\begin{aligned} pu &= qv \\ pu &= v(1-p) \qquad (\text{on replacing } q \text{ by } 1-p) \\ pu &= v - pv \\ v &= pu + pv \\ v &= p(u+v) \qquad (p \text{ is taken common}) \end{aligned}$$

Therefore,

$$\hat{p} = \frac{v}{u+v}$$

where, $\hat{p}$ is the equilibrium frequency of A. Similarly, we may replace p by $1-q$ and show that $\hat{q} = u/(u+v)$, where $\hat{q}$ is the equilibrium frequency of a allele.

Thus, when reverse mutation also takes place, the frequencies of the two alleles change till they reach mutational equilibrium frequencies. At equilibrium,

p = v/(u + v), *and there will be no further change in the values of* p *and* q *due to mutation.*

41.4. SELECTION

Selection is defined as differential rates of reproduction of different genotypes in a population. When the differential reproduction is the consequence of natural factors or the environment, it is called ***natural selection.*** But when it is brought about by human efforts, it is termed as ***artificial selection.*** Further, selection may operate on gametes or the haploid (n) phase (***gametic selection***), or it may act on zygotes or the diploid ($2n$) phase (***zygotic selection***).

41.4.1. Fitness

The ***rate of reproduction*** for a genotype is the average number of progeny produced by each individual of that genotype or by a mating among individuals of that genotype, and which contribute to the gametic pool of the next generation. Clearly, those progeny, which fail to reproduce are not considered while determining reproduction rate The reproduction rates of different genotypes are commonly expressed relative to that of the genotype having the highest reproduction rate; this relative value is called ***fitness, selective value*** or ***adaptive value*** and is denoted by W. The value of W is estimated as follows.

$$W = R_i/R_h$$

where, R_i is the rate of reproduction of the concerned genotype, while R_h is the rate of reproduction of that genotype, which has the highest reproduction rate. If the reproduction rates of the genotypes *AA, Aa* and *aa* were 16, 12 and 8, respectively, their fitness will be 1.0 (= 16/16), 0.75(= 12/16) and 0.5 (= 8/16), respectively. *The fitness may have any value between 0 and 1.*

The genotypes having fitness of 1.0 are favoured by selection, while those having fitness of less than 1.0 are selected against. The magnitude of selection against a genotype is called its ***selection coefficient*** or ***selection differential;*** it is denoted by s and is estimated as $1 - W$, *i.e.*, $s = 1 - W$. The values of s, like those of W, range between 0 and 1. Clearly, as the value of fitness of a genotype decreases, its selection coefficient increases, and *vice versa.*

41.4.2. Gametic Selection

When selection acts on gametes or on the haploid phase of life-cycle, it is called ***gametic selection.*** In most fungi, the main part of life-cycle is haploid; therefore, selection in these organisms will be gametic. In some higher organisms, which are normally diploid, gametes having certain genotypes are known to have differential survival or capability for effecting fertilization; such a phenomenon is called ***segregation distortion*** or ***meiotic drive***. There are only limited examples of meiotic drive, and it is often confined to only one of the two sexes.

For a gene with two alleles, there will be two genotypes for gametes, say, A and a, whose frequencies will be p and q, respectively. If a gametes have a selection coefficient of s, their fitness will be $1 - s$. There will be a decline in the frequency of a in the next generation that will equal sq. Since this will lead to a reduction in the population size, the value of q in the next generation will be $(q - sq)/(1 - sq)$ (Table 41.1). Thus *in each generation, the frequency of* a *allele will decline by the proportion* s *till it reaches zero, and* A *becomes fixed in the population.*

41.4.3. Zygotic Selection

Most higher organisms are diploid and their haploid phase (gametes) is only short-lived and totally dependent on the diploid phase. In these organisms, selection operates on the zygotes or the diploid phase, and is called ***zygotic selection.*** Zygotic selection may be, (1) against the recessive phenotype, (2) against the dominant phenotype or (3) in favour of the heterozygotes.

TABLE 41.1
Effect of gametic selection on gene frequency

Parameter	*Gamete genotype*		*Total*
	A	a	
Initial frequency	p	q	1
Selection coefficient	0	s	
Fitness	1	$1 - s$	
Frequency after selection	p	$q(1 - s)$	$1 - sq$*
Frequency in the next generation**	$p/(1 - sq)$	$(q - sq)/(1 - sq)$	

*$1 - sq = p + q\,(1 - s) = p + q - sq$ (remember, $p + q = 1$)

** The values p and $(q - sq)$ are divided by $1 - sq$ since it is the population size in the next generation (in place of 1).

41.4.4. Selection Against The Recessive Phenotype

Selection against the recessive allele acts on the recessive phenotype, say *aa.* It does not affect the heterozygotes *Aa* since the deleterious effects of a in them are masked by A. If the selection coefficient for *aa* zygotes is s, their frequency after selection will be $q^2(1 - s)$, *i.e.*, q^2W (where, $W = 1 - s$), and the frequency of allele a after selection will be $(q - sq^2)/(1 - sq^2)$. Therefore, Δq (or the rate of decline in q) will be $-spq^2/(1 - sq^2)$ (Table 41.2). Clearly, *as the value of* q *decreases, the value of* Δq *will also decline at an increasingly rapid rate since* Δq *depends on* q^2. If $s = 1$, $\Delta q = pq^2/(1 - q^2)$; if p is also very close to 1, the value of sq^2 will be close to zero (as q will be close to zero), and Δq will be $- spq^2$ or $- q^2$, which will also be close to zero. Therefore, *selection against rare recessive alleles is virtually ineffective in eliminating them. This is mainly because these alleles escape the effects of selection by being present mainly in heterozygotes.*

As long as q is less than 0.67, the frequency of *aa* ($= q^2$) will be lower than that of *Aa* ($= 2pq$); the lower the value of q, the greater will be the difference between $2pq$ and q^2. For example, if $q = 0.5$, $q^2 = 0.25$ and $2pq = 0.5$; if $q = 0.1$, $q^2 = 0.01$ and $2pq = 0.18$; if $q = 0.01$, $q^2 = 0.001$ and $2pq = 0.0198$. Thus, *as the recessive allele* a *becomes increasingly rare, a larger proportion of it will be present in the heterozygotes and, thereby, escape selection.* Consequently, the value of Δq becomes rather disappointing with low values of q. For example, if $s = 1$, and $q = 0.01$, it will require 900 generations for q to become 0.001, and it will be another 9,000 generations for it to further decline to 0.0001.

As a result, *Mendelian populations retain many recessive alleles whose fitness is less than 1, sometimes even zero. These deleterious recessives reduce the fitness of the population as a part of it is lost due to selection. The proportion of a population that is lost due to selection is called* ***genetic load.*** Thus, the larger the genetic load of a population, the lower will be its average fitness.

TABLE 41.2

Effect of selection against the recessive phenotype (or the recessive allele) on gene frequency

Parameter	*Genotype*			*Total*
	AA	Aa	aa	
Initial frequency	p^2	$2pq$	q^2	1
Selection coefficient	0	0	s	
Fitness	1	1	$1-s$	
Frequency after selection	p^2	$2pq$	$q^2(1-s)$	$1-sq^2$*

* $1 - sq^2 = p^2 + 2pq + q^2(1-s) = p^2 + 2pq + q^2 - sq^2$ (remember, $p^2 + 2pq + q^2 = 1$)

Frequency of allele a in the next generation (q_1) $= [pq + (q^2 - sq^2)]/(1 - sq^2)$

$= (q - sq^2)/(1 - sq^2)$ (since, $pq + q^2 = q$)

or $= [q(1 - sq)]/(1 - sq^2)$

and $\Delta q = q_1 - q$

$= -\dfrac{q - sq^2}{1 - sq^2} - q$ (on replacing q_1, by its value)

$= \dfrac{(q - sq^2) - [q(1 - sq^2)]}{1 - sq^2}$

$= \dfrac{q^2 - sq^2 - q + sq^3}{1 - sq^2}$

$= \dfrac{-sq^2 - sq^3}{1 - sq^2}$

$= \dfrac{-sq^2(1-q)}{1 - sq^2}$ (sq^2 is taken common)

Therefore, $\Delta q = spq^2/(1 - sq^2)$, since $1 - q = p$.

41.4.5. Selection Against the Dominant Phenotype

Selection against a completely dominant alleles acts on both homozygous (*AA*) and heterozygous (*Aa*) dominant phenotypes. As a result, the frequency of allele *A* will decline by a value of $-spq^2/[1 - sp(1 + q)]$ in each generation, where, s is the selection coefficient for *AA* and *Aa*, and p and q are the frequencies of *A* and *a* in that generation (Table 41.3).

TABLE 41.3

Effect of selection against the dominant phenotype on gene frequencies

Parameter	*Genotype*			*Total*
	AA	Aa	aa	
Initial frequency	p^2	$2pq$	q^2	1
Selection coefficient	s	s	0	
Fitness	$1-s$	$1-s$	1	
Frequency after selection	$p^2(1-s)$	$2pq(1-s)$	q^2	$1 - sp(1 + q)$*

* $1 - sp\ (1 + q) = p^2(1 - s) + 2pq(1 - s) + q^2 = p^2 - sp^2 + 2pq - 2spq + q^2 = p^2 + 2pq + q^2 - sp^2 - 2spq = 1 - sp^2 - 2spq$ (remember, $p^2 + 2pq + q^2 = 1$) $= 1 - sp(p + 2q) = 1 - sp(1 + q)$ (since, $p + 2q = p + q + q = 1 + q$)

The frequency of allele *A* in the next generation, $p_1 = \dfrac{p^2 - sp^2 + pq - spq}{1 - sp(1 + q)}$

$= (p^2 + pq - sp^2 - spq)/[1 - sp(1 + q)]$

$= (p - sp^2 - spq)/[1 - sp(1 + q)]$ (since $p^2 + pq = p$)

$= [p - sp(p + q)]/[1 - sp(1 + q)$

$= (p - sp)/[1 - sp\ (1 + q)]$ (since $p + q = 1$)

Therefore, $\Delta p = p_1 - p$

$= [(p - sp)/\{1 - sp(1 + q)\}] - p)$ (on replacing p_1 by its value)

$= \dfrac{(p - sp) - p[1 - sp(1 + q]}{1 - sp(1 + q)}$ [on multiplying p by $[1 - sp(1 + q)]$

$= \dfrac{(p - sp) - p[1 - sp - spq)]}{1 - sp(1 + q)}$

$= \dfrac{p - sp - p + sp^2 + sp^2q}{1 - sp(1 + q)} = \dfrac{-sp + sp^2 + sp^2q}{1 - sp(1 + q)}$

$= \dfrac{s(-p + p^2 + p^2q)}{1 - sp(1 + q)}$ (s is taken common)

$= \dfrac{s[-p + p(1 - q) + p(1 - q)q)]}{1 - sp(1 + q)}$

(p^2 rewritten as $p(1 - q)$ and p^2q as $p(1 - q)q$]

$= \dfrac{s[-p + p - pq + p(q - q^2)]}{1 - sp(1 + q)}$

$= \dfrac{s(-p + p - pq + pq - pq^2)}{1 - sp(1 + q)}$

$= -spq^2/[1 - sp(1 + q)]$ (since $-p$, $+p$ and $-pq$, $+pq$ cancel out)

In due course of time, A *will be eliminated and* a *will become fixed in the population. This happens because, unlike the recessive allele* a, *the dominant allele* A *is not protected from selection in the heterozygotes (*Aa*).*

Selection in Favour of Heterozygotes

- p and q change until equilibrium.
- Equilibrium value: $p = t/(s + t)$, $q = s/(s + t)$.
- $\Delta q = \dfrac{[pq\ (sp - tq)}{(1 - sp^2 - tq^2)}$
- Both alleles in equilibrium frequencies.
- A higher frequency of heterozygotes than expected.,

41.4.6. Selection in Favour of Heterozygotes

When selection favours the heterozygote *Aa*, both the homozygotes, *AA* and *aa*, are selected against. As a result, the frequencies of *A* and *a* alleles change in every generation, the Δq being $[pq(sp - t\,q)]/(1 - sp^2 - t\,q^2)$, where p and q are the frequencies of *A* and *a*, respectively, in a given generation, and s and t are the selection coefficients for the homozygotes *AA* and *aa*, respectively. *Ultimately,* q *will reach a value, the equilibrium value, when* sp = tq*; in this situation* Δq = 0. *The equilibrium will be attained when* q = s/(s + t) *or* p = t/(s + t) (Table 41.4). It is

TABLE 41.4
Effect of selection for heterozygotes on gene frequency.

Parameter	*Genotype*			*Total*
	AA	Aa	aa	
Initial frequency	p^2	$2pq$	q^2	1
Selection coefficient	s	0	t	
Fitness	$1-s$	1	$1-t$	
Frequency after selection	$p^2(1-s)$	$2pq$	$q^2(1-t)$	$1-sp^2-tq^2$*

$$*1-sp^2-tq^2 = p^2(1-s)+2pq+q^2(1-t) = p^2-sp^2+2pq+q^2-tq^2$$
$$= p^2+2pq+q^2-sp^2-tq^2$$
$$= 1-sp^2-tq^2 \quad (\text{since, } p^2+2pq+q^2=1)$$

Frequency of a after selection $(q_1) = (pq+q^2-tq^2)/(1-sp^2-tq^2)$

$$= (q-tq^2)/(1-sp^2-tq^2) \quad (\text{since, } pq+q^2=q)$$

Therefore, $\Delta q = q_1 - q$

$$= [(q-tq^2)/(1-sp^2-tq^2)] - q \quad (\text{on replacing } q_1 \text{ by its value})$$
$$= \frac{(q-tq^2)-q(1-sp^2-tq^2)}{1-sp^2-tq^2} \quad (\text{on multiplying } q \text{ by } 1-sq^2-tq^2)$$
$$= \frac{q-tq^2-q+sp^2q+tq^3}{1-sp^2-tq^2}$$
$$= \frac{-tq^2+sp^2q+tq^3}{1-sp^2-tq^2} \quad (q \text{ and } -q \text{ cancell out})$$
$$= \frac{sp^2q+tq^2(1-p)-tq^2}{1-sp^2-tq^2} \quad [tq^3 \text{ rewritten as } tq^2(1-q)]$$
$$= \frac{sp^2q+tq^2-tpq^2-tq^2}{1-sp^2-tq^2}$$
$$= \frac{sp^2q-tpq^2}{1-sp^2-tq^2} \quad (+tq^2 \text{ and } -tq^2 \text{ cancel out})$$
$$= \frac{pq(sp-tq)}{1-sp^2-tq^2} \quad (pq \text{ is taken common})$$

Thus q will become zero, if $sp - tq = 0$ or if $sp = tq$.

This will happen at equilibrium frequency of q $(= \hat{q})$.

Therefore, $\hat{q}$ is derived as follows. At equilibrium, $sp = tq$

$$sp + sq = tq + sq \quad (\text{add } sq \text{ to both the sides})$$
$$s(p+q) = q(s+t) \quad (s \text{ and } q \text{ taken common})$$
$$s = q(s+t) \quad (\text{since } p+q=1)$$

Therefore, $\hat{q} = s/(s+t)$

Similarly, we can show that, $\hat{p} = t/(s+t)$

[For example, $sp = tq$

$$sp + tp = tq + tp \quad (\text{Add } tp \text{ to both the sides})$$
$$p(s+t) = t(p+q) \quad (p \text{ and } t \text{ taken common})$$
$$p(s+t) = t \quad (\text{since } p+q=1)$$

Therefore, $\hat{p} = t/(s+t)$]

assumed that both s and t are constant and in the same direction generation after generation.

Obviously, the direction of Δq will depend on the value of q in a generation relative to the equilibrium value $\hat{q}$, which is based on the values of s and t. If q in a generation is below the value of $\hat{q}$, Δq will be positive, *i.e.*, q will increase due to selection for the heterozygotes. But the opposite will be the case if q in a generation were greater than $\hat{q}$.

Thus selection for heterozygotes will retain both the alleles A *and* a *in the population, and that these two alleles will be maintained in equilibrium. A symptom of selection for heterozygotes in a population is the greater than expected frequency of heterozygotes, while those of the two homozygotes are lower than expected.*

Random Drift
- Change in gene frequency due to chance or sampling error.
- Direction of change is random.
- Change larger in smaller populations.
- Genes become fixed; rate of fixation is $1/N$ per generation.

41.5. RANDOM DRIFT

Random fluctuation in gene frequencies of a population due to chance or sampling error is called ***random genetic drift, genetic drift*** or ***random drift.*** *As a rule, the change in gene frequency due to random drift does not occur in the same direction in all the generations; its direction may change from one generation to the next. Random drift is greater in a smaller population than in a larger one.* Wright estimated that even with a small amount of migration ($m = 0.0001$), fixation of alleles is likely to take place in many populations having the size of 5,000 in each generation; but if the population size was only 1,000, most of the populations would show fixation for one or the other allele. *In general, which allele of a gene becomes fixed in a population is a matter of chance.* If a large number of small populations are considered together, the proportion of populations showing fixation of A allele will be close to p and that having fixation of a will be approximately q, *i.e.*, the average frequencies of A and a, respectively, in the group of small populations.

Random drift is the consequence of sampling error. Since many natural populations are rather small, their each generation arises from a random sample of male and female gametes. In diploid organisms, gene frequencies in such samples would have the standard deviation of $\sqrt{pq/2N}$, where p and q are the frequencies of the two alleles for a gene, and N is the number of actual parents (individuals that participate in reproduction) of the population. Thus in a generation, gene frequencies may be expected to go up or decline by upto this value $(= \sqrt{pq/2N})$ purely due to chance in about 68% of the cases. Obviously, the value of standard deviation will become larger as the value of N decreases. Such fluctuations in gene frequencies may often lead to the fixation of an allele. The rate at which genes become fixed in a population is estimated by the formula, $1/2N$, and is called ***'rate of fixation'.***

The tendency toward fixation of an allele by random drift may be counteracted by selection. But if an allele were neutral, its fixation is postulated to depend mostly

on its mutation rate. The average number of generations necessary for the fixation of a new neutral mutation is estimated to be approximately four times the number of parents in each generation. The number of parents in a population is called ***effective population size (Ne)***. The effective population size differs from the actual population size due to the following reasons: (1) failure of some individuals to reproduce, (2) unequal number of males and females, (3) isolation of a population into several subpopulations, etc. When the population consists of male and female individuals.

$$Ne = 4Nf.Nm/(Nf + Nm)$$

where, *Nf* and *Nm* are the numbers of reproducing males and females, respectively. If $Nf = Nm$, $Ne = Nf + Nm$.

Several studies have shown that genetic drift has been an important factor in the variations that have developed among different populations. Wright concluded that under certain circumstances, random drift may produce notable changes in gene frequencies even in the face of selection or migration from other populations.

41.6. FOUNDER PRINCIPLE

A new population of a species is often initiated by a relatively small number of individuals of a population; these individuals are called ***'founders'*** of the new population. The gene frequencies and the chromosome arrangements present in the founders of a population may not be characteristic for the parent population, that is, they may differ from those of the parent population. Thus, the new population may acquire new gene frequencies and chromosome arrangements simply due to the fact that its founders possessed them. This genetic divergence created by the limited number of founders of new populations is called the ***founder principle.*** This genetic divergence among the populations may become enhanced as their environments may differ, and they may be subjected to natural selection in divergent directions.

It is suggested that in many cases the founder principle has been of considerable importance in race formation and speciation. It may be pointed out that founder principle is a form of random drift, and that in such populations random drift is likely to play a crucial role due to their small size.

QUESTIONS

1. List the various factors affecting gene frequencies in a population. Discuss the effects or migration and mutation on gene frequencies.
2. Define selection, fitness and selection coefficient. Discuss the effects of gametic selection and zygotic selection against the dominant phenotype on gene frequencies.
3. Discuss the effect of selection against the recessive phenotype on gene frequencies. Explain why selection against rare recessives is virtually ineffective.
4. Discuss the effect of selection in the favour of heterozygotes for a gene on the gene frequency.
5. Write short notes on the following:
   (*i*) evolutionary factors, (*ii*) migration,
   (*iii*) mutation, (*iv*) random drift,
   (*v*) founder principle, (*vi*) heterozygote advantage,
   (*vii*) fitness, (*viii*) selection coefficient,
   (*ix*) gametic selection, (*x*) zygotic selection,
   (*xi*) mutational equilibrium.

42

Evolution and Origin of Species

42.1. INTRODUCTION

A keen observer is impressed by the striking similarities and consistent differences among the various living forms; these differences and similarities form the basis of their classification into various taxonomic entities. But taxonomic classification assumes the various species to be 'fixed' entities in that each species has specific morphological features, which were and are expected to remain unchanged. This view places the sole emphasis on the differences among various species, and ignores their similarities. However, attempts to explain the similarities in structures and functions of various living forms date back to early Greeks and Romans. All these attempts culminated in the concept of ***'organic evolution'***, which envisages that all the living forms of today developed from a common ancestor. That is, the various life forms are related by descent, which accounts for the similarities among them. The idea of evolution envisages the 'species' to be fluid rather than fixed; it further considers them to be constantly undergoing changes to yield more and new species.

The idea of organic evolution was not widely accepted until 1859 when Darwin published his classic work *'The Origin of Species.'* This work contained a large body of evidence in favour of the idea that evolution had taken place and was still taking place, as well as an attractive hypothesis to explain the mode of evolution. Subsequently, various concepts regarding the mechanism of evolution were developed by Haldane, Fisher, Wright and several others. Informations from diverse areas of study, *e.g.*, geology, paleontology, taxonomy, population genetics, biochemistry, molecular genetics, etc., have been sifted and resynthesized in a quest to understand evolution. It is generally believed that evolution has occurred and is occuring at present, but its precise mechanics is not understood and is a subject of debate and continuing investigations. Some of the basic aspects of evolution and the basic features of the evolutionary process will be briefly described here; for a more detailed discussion the reader is referred to one of the books devoted solely to evolution.

42.2. THEORIES OF EVOLUTION

Many scientists have conjectured on how and why evolution took place; these are typified by the following four theories: (*i*) Lamarckism, (*ii*) Darwinism, (*iii*) mutation theory and (*iv*) neo-Darwinism.

42.2.1. Lamarckism

This theory was propounded by Jean Baptiste Lamarck (1744-1829), a French biologist, through his numerous writings. There are four postulates of this theory as listed below.

Theories of Evolution

- Lamarckism (Lamarck, 1744-1829)
- Darwinism (Darwin, 1859)
- Mutation theory (de Vries, 1840-1935)
- Synthetic theory (several contributors)
  - — Gradualistic model (Darwinian view)
  - — Punctuational Model (Goldschmidt and others)

1. Living organisms and their organs tend to increase in size during evolution.
2. New organs would develop in an organism if this development was needed for its survival.
3. Those organs that are in frequent use will develop more and more as their constant use will make them more developed, while the organs not being used will tend to become weaker.
4. The modifications thus produced due to the use and disuse in the traits of an organism will be inherited and would accumulate with time.

The essence of this theory may be explained with the help of following example. Giraffes feed on tree leaves. As the easily accessible leaves are exhausted, they will have to stretch their necks higher and higher in order to feed. This stretching will cause some increase in the length of their necks, which will be passed on to the next generation. In each generation, therefore, their necks will be subjected to this type of stretching; its effect will accumulate over generations giving rise to the present day long necks of giraffes.

The postulates of Lamarck were criticised even during his lifetime, and several experiments showed the 'theory of inheritance of acquired characters' to be incorrect. Many followers of Lamarck introduced various modifications of his theory; one of the important variations being that changes in the traits of organisms were induced by the environment (in place of the need postulated by Lamarck). However, *the theory is now only of academic interest without any followers.*

42.2.2. Darwinism

The theory of origin of species through *natural selection* was proposed by Charles Darwin (1809-1882), an English biologist; his theory is fully elaborated in his book *"The Origin of Species"* first published in 1859. Darwin visited several islands of the Atlantic Ocean near the coast of south America, and of south Pacific Ocean. He collected a vast amount of data on the biological entities in these islands; these were brilliantly used to develop his arguments on evolution. These arguments may be simply stated as follows.

1. Normal reproduction by the members of a species increases their number at a geometric rate. However, there are only limited food supplies and other environmental opportunities.
2. As a result, members of the same species compete with each other for survival, which Darwin called, the ***struggle for existence.***
3. Members of a single species show variation for various traits, and Darwin assumed these variations to be hereditary. He suggested that the 'fittest' members of the species will 'survive' the 'struggle for existence'; this he termed as ***the survival of the fittest',*** and the factor responsible for this phenomenon was called ***natural selection.***

Darwin, along with A.R. Wallace, concluded that *natural selection coupled with a change in the environment in which a species lives will bring about the evolution from it of another species different from the parental species.*

Thus the two main features of the Darwinian theory of evolution are: (i) origin of hereditary variations in a species, and (ii) selective multiplication of those variations, which make the species more adapted to the prevailing environment through natural selection. According to this scheme, natural selection is the force, which determines the direction in which evolution in a species would proceed, and the direction of this influence will depend on the prevailing environment. *Thus natural selection is central to evolution, and this constitutes the Darwinian theory of evolution.* Natural selection can operate only when genetic variation exists in a population (individuals of a single species). Darwin recognised two types of variation in living organisms: (*i*) continuous (quantitative in nature) and (*ii*) discontinuous (discrete variations arising due to, what Darwin called, *sports* or mutations; akin to qualitative variation).

Darwin believed that continuous variation was the basis for evolution, but he had no clue as to how the continuous genetic variation arose. He attempted to explain its origin on the basis of a modification of the theory of inheritance of acquired characters (see, Chapter 1). Many scientists have endeavoured to demonstrated the existence of considerable genetic variation within species in order to elucidate the modes of origin of these variations and to analyse the various aspects of selection on the variation so generated. Sometimes, 'Darwinism' is used as a synonym of 'evolution', which is not correct: evolution, as stated earlier, is the process of origin of more complex life forms from simpler ones, while Darwinism is a theory (of natural selection) of how evolution came by.

42.2.3. Mutation Theory

This theory was put forth by Huge de Vries (1840-1935), a Dutch botanist, who described a large number of discrete variations in *Oenothera lamarckiana.* He coined the term 'mutation' for the phenomenon of appearance of sudden heritable changes; the various mutants were called new and separate species by him. *He suggested that new species could arise in a single step due to mutation and the natural selection was not important for such an evolution. Further, since mutations are random, it was assumed that evolution was also random,* i.e., *it did not proceed in a direction.*

One of the important consequences of natural selection is that populations are expected to become progressively more and more adapted to their environment. Mutationists considered that populations were preadapted and that adaptations did not arise due to natural selection. *It is now universally accepted that mutations are the ultimate source of all genetic variation present in the biological world, but that natural selection is required to bring about the evolution of new living forms.*

42.2.4. Neo-Darwinism

The contribution of numerous scientists after Darwin have been synthesized to develop a cohesive and plausible theory of the evolutionary process; therefore, it is often referred to as ***'synthetic theory'***. Some of the main contributors to this theory

are, Fisher, Dobzhanski, Haldane, Wright, Meyer, Stebbins, Lewontin and others. These workers have attempted to explain (1) the mode of origin of genetic variation, (2) the origin and role of reproductive isolation, (3) the occurrence and role of natural selection, and (4) the identification and elucidation of the effects of other evolutionary forces, *e.g.,* random drift, molecular drive, etc. Sewall Wright proposed the ***shifting balance theory of evolution*** which perhaps, is the most far-reaching; it encompasses, polymorphism, pleiotropy, the relation between genotype and fitness and population structure, and, more particularly, population size.

42.3. SUCCESSION AND FOSSIL RECORD

The remains of organisms embedded in rocks are known as ***fossils.*** The approximate period of time when a given fossil arose can be reliably determined through several procedures, carbon-dating being one of the common techniques. ***Carbon dating*** is based on the level of radioactive carbon detected in a fossil, which is then used for working out its period of origin. An examination of the available fossil depicts the following general picture: *progressively more and more complex life forms have successively appeared beginning with the simplest ones. For example, the available evidence shows the order of appearance to be bacteria, sponges, fish, amphibia, reptiles, birds and mammals.*

This clearly shows the succession of life forms, but it does not demonstrate that they arose from the previously existing forms. The fossil records are sporadic and incomplete; they have large gaps associated with the periods when large physiological changes in the life forms are expected to have taken place. Thus fossil records do not have the intermediate links between the major groups of organisms, *e.g.,* the links between fish, amphibians, reptiles, birds and mammals. Several arguments have been advanced to explain these gaps. One of the arguments is as follows: the intermediate life forms linking two major groups would have been rather few in number since they would be relatively unadapted as they were still evolving. Therefore, the chances of such forms being fossilized would be rather remote. Another, more accepted, argument is based on the fact that the origin of major taxonomic groups, such as fish, amphibia, reptiles and mammals, has coincided with massive global upheavals called ***geological upheavals.*** These geological upheavals may confuse the fossil record making it difficult to detect the relationships between different major groups.

42.4. GRADUALISTIC MODEL OF EVOLUTION

According to the Darwinian theory, natural selection acts on continuous genetic variation present in a species and favours an increase in the numbers of those individuals, which possess the features that confer on them a greater adaptation to the prevailing environment. Thus in every generation, there will be selection for the character combinations possessing improved adaptation from among the variants present in a population. This will tend to gradually accumulate more and more changes in the characters of the members of that species, and in the due course of time they will become divergent from the original species. It is assumed that this population is subjected to a changed environment from that of the original species. Thus in Darwinism, evolution is perceived to be a gradual process of accumulation of genetic changes leading to the formation of new species (***speciation***). This

concept of continuous evolution at a slow rate is called ***gradualism***, and the phenomenon is termed as ***microevolution.***

Thus adaptations are expected to arise due to natural selection, and, conversely, they would disappear if the selection is removed. This seems to be the case in, at least, experimental populations. Dobzhansky and coworkers maintained seven strains of *Drosophila pseudoobscura,* each of which was homozygous for some deleterious genes, in crowded cultures for 50 generations. In five of these strains, survival showed a marked increase at the end of the experiment; irradiation of the flies with X-rays elicited this response in six of the seven cultures. Clearly, these strains became more adapted to the new environment due to a gradual accumulation of favourable mutations/gene combinations over a period of 50 generations.

The converse of this, *i.e.,* a loss of adaptation in the absence of selection, is evident from the findings of another experiment by Carson and associates published in 1982. They isolated a parthenogenetic female strain of *Drosophila mercatorum* from the normal population through a selection for parthenogenesis. This strain has been maintained parthenogenetically for 20 years without any contact with males, since mating was not necessary for reproduction. These female flies have also lost their ability for mating. These findings have been interpreted as follows: the mating behaviour of males and females is genetically determined and develops due to an interaction between the two sexes. The maintenance of females in isolation from males removed the selection for mating ability. The occurrence of random mutations has, therefore, impaired the mating ability of these parthenogenetic females.

42.5. PUNCTUATIONAL MODEL OF EVOLUTION

Although there is a considerable evidence in favour of microevolution, it is unable to explain the origin of major changes during evolution, called macroevolution. ***Macroevolution*** includes the processes that lead to the formation of new species, genera, families, orders and higher taxonomic categories. It may involve phenomena like hybridization and polyploidy in plants and major chromosome changes both in plants and animals. Whether the process of macroevolution is the same as that of microevolution, *i.e.,* gradual accumulation of genetic changes, through natural selection, or it is quite distinct is not clear. A theory of sudden and rapid changes leading to macroevolution was proposed by R. Goldschmidt (1878-1958) during 1940's; this theory has been since developed into the ***'punctuational model'*** of evolution.

According to this model, large scale abrupt changes in organisms have taken place over relatively short geological periods; these changes account for the appearance of new large taxonomic groups such as orders and phyla. These abrupt changes take place when a population is subjected to a drastically changed environment. Most individuals of such a population will die, but some variants may survive the hostile environment. These variants would undergo genetic changes due to gene mutations, changes in chromosome number and structure, etc. Consequently, they would undergo a rapid change in their features in the face of extreme selection pressure imposed on them by the environment.

Thus the punctuational model has all the essential features of the gradualistic model, but it places a far greater emphasis on the role of random genetic drift. Further, the effect of selection in this model is more extreme and less refined than

that in the gradualistic model. The small population size assumed in this model has the following two important consequences.

1. Random mating among individuals of the population spreads the new genes throughout the population rather rapidly.
2. Chance is more important than natural selection in bringing about a change in the population.

Fixation of Mutant Alleles
- Selection
- Random drift
- Molecular derive

42.6. FIXATION OF MUTANT ALLELES

Speciation (formation of new species) involves the divergence in characters of the different populations of a species from those of the parental species. Since these differences in the characters are inherited, the new species acquires some new alleles/genes as compared to the parental species. These new alleles/genes would originate through mutation (*new alleles*) or through gene duplication coupled with mutation (*new genes*). But once new alleles arise, these have to replace the existing alleles in all the individuals of the population so that the new species is relatively homogeneous. There is considerable controversy as to how this replacement occurs; there are three different mechanisms, which may achieve such a replacement: (1) selection, (2) random drift and (3) molecular drive.

42.6.1. Selection

Selection is defined as differential rates of reproduction of various genotypes. The rate of reproduction of a genotype is compared with that of the genotype having the highest reproduction rate (see, Chapter, 41). The selection leads to a reduction in the number of individuals present in the next generation as compared to when selection were absent; the proportion of individuals of a population lost due to selection is called ***genetic load.*** Genetic load may be either due to physical death (actual death) of individuals or due to their genetic death. ***Genetic Death*** refers to the inability of an individual to reproduce due to sterility, inability to find mate, etc. All natural populations carry many lethal, sublethal and subvital genes. Since natural populations are adapted to their environment, most of the mutations occurring in them would be detrimental. Further, mutations generate the genetic variation necessary for the operation of natural selection. Therefore, all natural populations pay the penalty of genetic load for retaining genetic variation.

Genetic load of a population may originate due to mutation producing inferior alleles (***mutational genetic load***), or it may be due to segregation generating inferior genotypes (***segregational genetic load***). *Natural selection would tend to favour a beneficial mutant allele causing a gradual increase in its frequency, perhaps, till it replaces its nonmutant allele in the population,* i.e., *till the fixation of the mutant allele. Such a selection would, however, impose a considerable genetic load on the species. If selection is operating on several loci, the genetic load may become so great as to cause the extinction of the species.* For example, if the favoured allele is recessive, approximately $30N$ (N = number of individuals in the population)

individuals must be lost for this allele to replace its normal allele; this estimate is irrespective of the phenotypic effect involved. If the population is assumed to sacrifice 10% of its individuals, the replacement of a single allele would require 300 generations ($= 30N \times N/10$). Kimura has estimated that to account for all the amino acid substitutions in its proteins, each vertebrate species must have undergone at least one complete amino acid substitution per year. This would mean that each vertebrate species must reproduce at least $90N$ individuals in each generation (assuming one generation to be of 3 years) to escape extinction if selection were the sole cause of gene replacements.

'Selectionists' (scientists considering selection as the sole cause of gene replacements) have attempted to propose selection schemes that confer minimal genetic load on the populations, *e.g.,* frequency-dependent selection. They have also attempted to resolve the problem of genetic load due to selection by proposing the concepts of "threshold" numbers of polymorphic loci in a population, replacement of linkage blocks in place of individual genes, etc.; all these concepts seek to minimise the genetic loads of populations due to natural selection. Selectionists emphasize the following findings to stress the importance of selection in evolution.

1. In many cases, an association between particular enzyme alleles and particular environmental conditions have been found, *e.g.,* distribution of sickle-cell anaemia gene and malaria. This would indicate that the different alleles were maintained through selection.
2. Several other cases of enzyme polymorphism, although their advantages are not immediately clear, are difficult to explain on a purely random basis. For example, the allozyme patterns of five different enzymes were broadly similar in the *E. coli* samples isolated from intestinal tracts of lizards and humans, and from localities as widespread as New Guinea and Iowa. These observations can be hardly explained on the basis of neutral mutations and random drift.
3. In several cases, an association may exist between the function of an enzyme and the degree of polymorphism that it shows. For example, enzymes with more restricted uses show significantly less polymorphism than do enzymes whose substrates are more variable.

42.6.2. Random Drift

In view of the high genetic load due to selection against deleterious mutant alleles, Kimura and coworkers proposed that the most amino acid changes in proteins are neutral in their effect, *i.e.,* they do not affect the survival of the organisms carrying them. Since there will be no selection either in favour of or against such alleles, their fixation will occur due to recurrent mutation and random genetic drift; this fixation will not be accompanied with any genetic load. ***Random genetic drift*** is the chance fluctuation in gene frequencies of small random mating populations. This hypothesis is called the ***'neutralist hypothesis'*** or, sometimes, ***'non-Darwinian hypothesis of evolution'.***

The neutralist hypothesis is supported by the occurrence of large amounts of protein and enzyme polymorphisms in almost all the species studied; the estimates range from 20-30% to 60-70% of the proteins being polymorphic. Even if it is assumed that these polymorphic alleles will never be fixed, the maintenance of polymorphism itself at so many loci would incur an enormous and intolerable

genetic load. For example, the maintenance of polymorphism at 500 loci by means of heterozygote superiority would reduce the fitness of a population to 0.007, *i.e.*, only 7 out of 10,000 individuals will survive and reproduce. This is the case when the selection differential against the two homozygotes is only 0.02, that is a 2% reduction in their fitness.

Thus the evidence supporting the neutralist hypothesis is quite convincing. But in some cases, at least, selection is definitely responsible for the maintenance of polymorphism. However, the selectionist and neutralist models need not be mutually exclusive. Further, some mutations may be neutral under certain environments, but they may have selective value under other conditions.

42.6.3. Molecular Drive

Some novel molecular mechanisms may lead to a rapid replacement of some nucleotide sequences by other sequences. The fixation of genetic variants in a population through such molecular processes is called ***molecular drive.*** The three known mechanisms of molecular drive are: (1) unequal chromatid or chromosome exchange, (2) gene conversion, and (3) gain or loss of DNA through transposition.

Thus families of genes and noncoding sequences within a population may be fixed as a consequence of molecular drive. This drive may be both random and directional, and, more importantly, would not incur a genetic load by itself. The unusual patterns of gene fixation due to molecular drive are proposed to permit the origin of biological novelty and species discontinuity, which are not accomplished by selection or genetic drift.

42.7. GENETIC VARIATION

The extent of genetic variation present in a population can be estimated at several levels, some of which are as follows: (1) phenotype, (2) chromosomes (generally, rearrangements), (3) proteins and enzymes (electrophoretic variation), (4) amino acid sequences of specific proteins, (5) nucleotide sequences of specific genes, and (6) nucleotide sequences of the entire genomes (this being the ultimate measure of variation). A brief discussion on variation estimated in terms of phenotypic expression, and amino acid and nucleotide sequences is given below.

42.7.1. Phenotypic Variation

Although members of a species show remarkable similarity in their phenotypes, they carry a large variety of gene mutations. For example, it was estimated that only 17% of the flies in natural populations of *Drosophila persimilis* are free of lethal mutations. Another indication on concealed genetic variability in natural populations is the existence of various chromosome rearrangements; these are widespread in many species, *e.g.*, grasshoppers, snails, *Oenothera, Datura, Drosophila*, etc. A much larger amount of genetic variability is indicated from the studies on electrophoretic patterns of various proteins and enzymes. It has been estimated that various species ranging from man to *Drosophila* may exhibit polymorphism for 28% (man) to 81% (*Drosophila willistoni*) of their proteins.

The remarkable uniformity for various characters of the members of a species is proposed to be due to the canalization of development of those characters that are associated with fitness. ***Canalization*** is the phenomenon by which the development of a character is normally unaffected by the common variations in genotype as well

as environment. Canalization is disrupted by environmental shocks and by some mutant genes. Further, genotypes with increased or decreased canalization may be selected; such a selection is called ***canalization selection.*** Thus the canalization of development of characters will conceal a considerable amount of genetic variability for the trait since this variability will not be expressed at the phenotypic level (see, Chapter 15).

Another phenomenon that tends to conceal and preserve genetic variability has been termed as genetic homoeostasis by Lerner. ***Genetic homoeostasis*** is the property of a population to resist the efforts (through selection and/or inbreeding) to reduce its genetic variability beyond a certain level. Genetic homoeostasis is proposed to arise due to the particular gene frequencies established for a population over a long period through natural selection. Obviously, these gene frequencies confer the optimum fitness on the populations. Therefore, a rapid departure from them due to selection would reduce the fitness of these populations, thereby resisting the change imposed. Another mechanism for the origin of genetic homoeostasis, according to Lerner, may arise from the necessity for the maintenance of certain levels of heterozygosity to ensure normal development.

42.7.2. Amino Acid Sequences of Specific Proteins

Proteins are basic to the development of various characters. Therefore, phenotypic divergence among species must be based in the diversity of their proteins. Amino acid sequences of several different proteins from different taxa have been compared to assess their relatedness. One such protein is the haemoglobin, which is proposed to have originated from the muscle protein myoglobin. A comparison of the amino acid sequences of α, β, γ and δ polypeptides of human haemoglobin indicates that they all originated from the same ancestral protein (Fig. 42.1). The α and β chains were the earliest to differentiate (differences in 52% of residues), while the β and δ chains appear to have differentiated more recently (differences in 7% of residues).

The rationale for these comparisons is simple; a high degree of similarity in the amino acid sequences of two proteins cannot be coincidental but is indicative of a common lineage. Further, fewer the differences in amino acid sequences of two proteins, the more recently in the evolutionary history they must have diverged. The genes for the four human haemoglobin chains are presumed to have originated through gene duplication, followed by accumulation of mutations leading to their differentiation.

When haemoglobins from different species are compared, the number of differences in their amino acid residues usually compare well with their already deduced phylogenetic relationships. The alpha and beta chains of the same species generally show more differences than do the alpha or the beta chains of different species. But in many instances, the inferences drawn from comparisons of a single protein may be highly misleading. For example, haemoglobin of man and gorilla differ by a single amino acid, which is the extent of the difference exhibited by many mutant human haemoglobins.

Cytochrome C is another extensively investigated protein. Cytochromes are involved in oxidative phosphorylation. Cytochrome C has 104 amino acid residues. A comparison of amino acid sequences of cytochrome C from different organisms shows that, as in the case of haemoglobin, the degree of their relatedness so

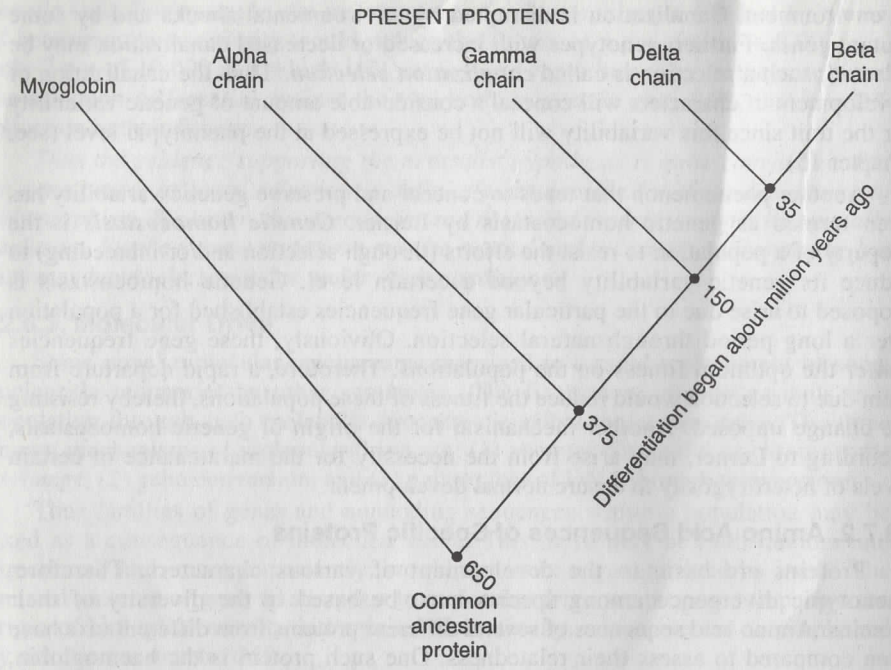


Fig. 42.1. A postulated phylogenic relationship among myoglobin and α, β, γ and δ polypeptides of haemoglobin based on the similarities and differences in their amino acid sequences.

estimated agrees well with their phylogenetic relationships (Fig. 42.2). For example, human and tuna fish cytochrome C show 12-13 amino acid differences, while those from man and yeast differ by 46-48 residues.

42.7.3. Nucleotide Sequences

Any genetic change must be reflected in the DNA base sequence. The analysis of nucleotide sequence of the complete genome has become feasible even in the case of eukaryotes. In addition, a considerable information on nucleotide sequence homology between DNA from two different species, say X and Y, can be obtained through DNA hybridization. In such studies, the total DNA from one species, say X, is radioa-ctively labelled. The two DNA samples (designated as X and Y) are denatured, mixed and allowed to renature. Preferential formation of labelled X-Y heteroduplexes is ensured by using a small amount of X relative to Y. The X–Y hybrid DNA is easily identified since it is radioactive; these heteroduplexes are isolated and their melting point is determined. It has been found that for each 1% difference in nucleotide sequence of X and Y DNA, the melting point of their hybrid (X:Y DNA) molecules is reduced by 1.6°C.

Amino acid sequences of haemoglobin, cytochrome C and fibrinopeptides, and the nucleotide sequences of the total DNA from cow, sheep and pig have been compared. The extent of differences, among cow-sheep and cow-pig comparisons

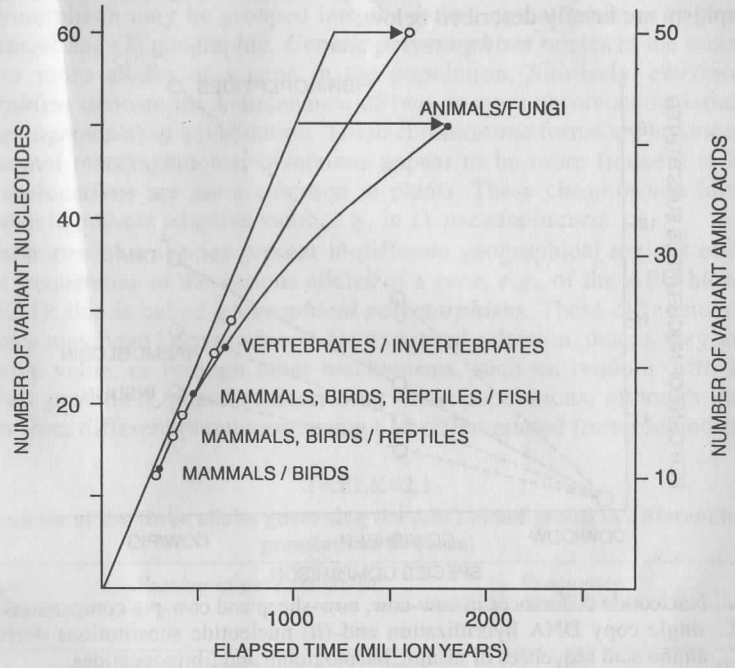


Fig. 42.2. The relationship between the numbers of variant amino acids in the cytochrome C (•—•) or the numbers of variant nucleotides in the corresponding genes (○—○) of different classes and phyla of organisms and the time elapsed since the divergence of the corresponding lines of the evolutionary descent. The straight line on the extreme left applies to both amino acid and nucleotide differences but is not corrected for multiple changes in a single residue, while the remaining two curves are appropriately corrected.

obtained from the four macromolecules agreed closely; they showed sheep to be closer than pig to cow (Fig. 42.3).

Genetic polymorphism
Polymorphism maintained by:
- Frequency-dependent selection
- Heterozygote superiority
- Cyclical selection

42.8. GENETIC POLYMORPHISM

The maintenance of two or more alleles of a gene in a Mendelian population exposed to the same environment throughout is called ***polymorphism.*** *In a constant environment, polymorphism is maintained due to an equilibrium between the forces of selection, mutation, migration and, possibly meiotic and molecular drives. Even if a balance between these evolutionary forces is lacking, polymorphism may be maintained by several other mechanisms. Such polymorphisms are known as*

balanced polymorphism. Some of the mechanisms that could maintain polymorphism are briefly described below.

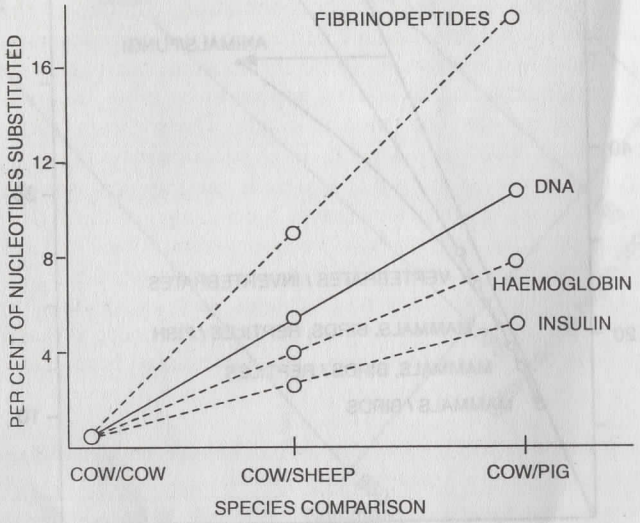


Fig. 42.3. Nucleotide differences in cow-cow, cow-sheep and cow-pig comparisons using (*i*) single copy DNA hybridization and (*ii*) nucleotide substitutions derived from amino acid sequences of insulin, haemoglobin and fibrinopeptides.

42.8.1. Frequency-Dependent Selection

In ***frequency-dependent selection,*** the frequency of an allele determines the fitness of the individuals expressing it. *In such a case, the lower is the frequency of an allele, the higher is the fitness of the genotypes expressing it. Such a selection would select for the less frequent and select against the more frequent allele, irrespective of which of the two alleles is more or less frequent.*

42.8.2. Heterozygote Superiority

Another mechanism for the preservation of polymorphism is ***heterozygote superiority.*** A selection for the heterozygotes at a locus will retain both the alleles of a gene in the population at equilibrium frequencies of $q = t/(t + s)$, where q is the equilibrium frequency of allele *a,* and t and s are the selection differentials for the homozygotes *AA* and *aa,* respectively (see, Chapter 41). This mechanism has been shown to be operative at several loci, the sickle-cell anaemia gene being one of the more documented ones.

42.8.3. Cyclical Selection

A third means of perpetuating balanced polymorphism involves the selection for different genotypes at a locus in different seasons or different years due to a fluctuating environment; this is termed as ***cyclical selection*** (Section 42.10). Consequently, one allele will be selected for at one point of time, while another will be favoured during another period, thus maintaining all the alleles of a gene in the population. Cyclical selection has been demonstrated for chromosomal

rearrangements in *Drosophila pseudoobscura* by Dobzhansky and coworkers, and in several other cases.

Polymorphism may be grouped into the following categories: (1) genetic, (2) chromosomal and (3) geographic. ***Genetic polymorphism*** relates to the maintenance of two or more alleles of a gene in the population. Similarly, ***chromosomal polymorphism*** denotes the maintenance of two or more chromosome structures or forms (arrangements) in a population. These chromosome forms are produced due to chromosomal rearrangements; inversions appear to be more frequent in animals, while translocations are more common in plants. These chromosome forms have been shown to possess adaptive value, *e.g.,* in *D. pseudoobscura.*

Populations of a species present in different geographical regions often show different frequencies of the various alleles of a gene, *e.g.,* of the ABO blood group (Table 42.1); this is called ***geographical polymorphism.*** These differences in gene frequencies may have been produced due to natural selection, that is, they may have an adaptive value, or through other mechanisms, such as, random drift. In many cases, they provide a means of describing these populations, although individual members from different populations cannot be differentiated from each other on this basis.

TABLE 42.1

Frequencies of the three alleles governing the ABO blood group in different human populations (= races)

Population	*Number of persons tested*	*Frequency*		
		i	I^A	I^B
Americans (white)	20,000	0.67	0.26	0.07
English	4,032	0.71	0.24	0.06
French	10,433	0.64	0.30	0.06
Indians	2,357	0.55	0.18	0.26
Chinese	2,127	0.59	0.22	0.20
Japanese	29,799	0.55	0.28	0.17
American Indians				
Navajo tribe	359	0.87	0.13	0.00
Black tribe	115	0.49	0.51	0.00
Australian aboriginals	243	0.69	0.31	0.00

42.9. Sources of Genetic Variation

All genetic variation is ultimately ascribable to ***mutation.*** But mutation produces only new alleles of existing genes; it cannot create new genes. New genes originate through ***duplication, followed by recurrent mutation*** in the duplicate gene (see, Chapter 7). Integration of transposable elements and episomes into the genome of an organism would also generate some genetic variation. A phenomenon called ***'gene fusion'*** produces ***'hybrid genes'*** from tandem duplications. One of the mechanisms for gene fusion is unequal crossing over, the same process which generates tandem duplication. One such instance of gene fusion is known to affect the genes for δ and β chains of human haemoglobin. Unequal crossing over results in the fusion of these genes producing hybrid δβ and βδ genes. The hybrid gene δβ (the δ sequence at the amino end and the β sequence at the carboxyl end of polypeptide) produces the phenotype called 'Hb-Lepore', while individuals having

the hybrid gene βδ (the β sequence at the amino end and the δ sequence at the carboxyl end of polypeptide) are known as 'anti-Lepore'. Thus gene fusion, if tolerated by the organism, may generate novel genes.

The existing alleles of various genes are sorted out in numerous combinations by segregation and recombination, which generates further variability. ***Segregation*** releases variability stored in heterozygotes, while ***recombination*** does the same when linkage, particularly repulsion phase linkage, is involved. Another mechanism for generating genetic variability is ***hybridization*** between different races and species. A form of hybridization, called ***introgressive hybridization,*** involves the transfer of a few genes or gene blocks from one species into the other. Introgression of genes from *Tripsacum* is considered to have played an important role in the evolution of maize.

Polyploidy, both auto- and allo-, also generates genetic variation; this mechanism has played an important role in the evolution of many plant and animal species (see, Chapter 9). ***Structural chromosome changes*** may sometimes bring about a change in chromosome number, but they often affect chromosome morphology. Some of them, *e.g.,* inversion, reduce or prevent recombination between the genes located in the inverted segment, thus facilitating the accumulation of genetic variability in such segments. Further, translocations introduce a novel feature of partial male sterility in translocation heterozygotes. *All these features of chromosomal rearrangements facilitate race and species differentiation.*

42.10. NATURAL SELECTION

The chief contribution of Darwin was the identification of natural selection as the force directing organic evolution. *Natural selection is the differential reproduction of various genotypes of a natural population in response to the natural environment in which the population exists.* Clearly, this concept of natural selection differs from the original Darwinian concept, which assumed the natural selection to weed out unfit genotypes from populations, most likely by causing their death. But the modern-day concept of natural selection perceives it as reducing the rates of reproduction of unfit genotypes rather than causing their elimination by death. It may also be emphasized that selection does not act on genes or alleles themselves; it acts on the phenotypes produced by them. Similarly, selection seldom operates on individual genes; it does so on gene combinations present in the different individuals of a population. In other words, *selection generally acts on the total phenotype of individuals and not on the phenotypes produced by individual genes. In addition, a gene may be detrimental in one genetic background, while it may be beneficial in another one. In other words, the selective value of a gene changes with the environment.*

42.10.1. Fishers' Fundamental Theorem of Natural Selection

In response to natural selection, the fitness of a population increases at a rate that is proportional to the genetic variability for fitness present in the population; this is known as the ***Fisher's fundamental theorem of natural selection.*** Since every population is subjected to a changing environment over a long period of time, it is crucial that populations retain genetic variability in order to survive. We have seen that the maintenance of variability imposes a genetic load on populations, which is the cost a population must pay for its long-term survival.

42.10.2. The Concept of Optimum Phenotypes

Every population may be considered to have evolved through a long period of natural selection. Therefore, it may be expected to have achieved an optimum level of adaptation to the prevailing environment. This will be reflected in most of their individuals achieving phenotypes that are optimally adapted to their environments, *i.e.,* many phenotypes will tend to cluster around the value that has the highest fitness;' this value, of necessity, will be the *population mean.* A corollary of this consideration is that phenotypes departing, on either side, from this optimal value will have a lower than optimal fitness, that is, they will be selected against. This concept of optimal phenotype has been found to hold good in many studies. For example, the birth weights of 6,693 female babies born in a London hospital during a certain period ranged from about 2.5 to 10 lbs. Of these, 274 babies died during the first month. The eight-pound class showed the lowest mortality; babies having higher or lower birth weights than this showed markedly higher mortality rates. *Thus natural selection ordinarily tends to favour those phenotypes that are optimally adapted to their environment and it acts against phenotypes having higher or lower values than this phenotype.*

42.10.3. Types of Selection

Selection may be grouped into the following five categories depending on the type of phenotypes selected for: (1) stabilizing selection, (2) directional selection, (3) cyclical selection, (4) disruptive selection and (5) frequency-dependent selection.

Stabilizing selection selects for intermediate phenotypes, and acts against the extreme phenotypes, *e.g.,* birth weight in humans. A selection for one of the extreme phenotypes is called ***directional selection.*** The role of directional selection in evolution is also of importance when the environment of a population is changing, and only the extreme phenotypes happen to be adapted to the new conditions. When selection favours one phenotype in one generation, and another one in the next generation, it is called ***cyclical selection.*** Cyclical selection is likely to be operative when the environmental conditions fluctuate from one generation to the next. In the case of ***disruptive selection,*** both the extreme higher and lower phenotypes are selected for in the same generation. Such a selection is expected to operate when parts of a single population are subjected to two diverse environments in the same generation. Disruptive selection will cause the two parts of a population to diverge from each other. The selective value of different alleles of a gene increases as their frequency decreases, and *vice-versa,* in the case of ***frequency-dependent selection;*** in some cases, natural selection may be of this type.

42.11. DIFFERENTIATION OF RACES

Races are different populations of a species that differ in terms of gene frequencies or chromosome rearrangements. Since all the individuals of a population share the same gene pool, each race of a species has a separate gene pool. It implies that there is little or no gene-flow between different races of a species since migration among races would tend to minimise the differences between them. However, when members of different races of a species are brought together, they interbreed freely and there is no reproductive isolation (see, later) between them.

An example of different races of a single species is the races of *Homo sapiens* (man). This species is generally classified into five main races. *European group,*

includes a variety of populations known as the white caucasians. The African blacks form the *African group,* while the *Asian group* consists of the Mongoloid peoples and the populations in the India-Pakistan subcontinent. All the aboriginal populations found on the American subcontinent are placed in the *American group.* Similarly, the *Pacific group* consists of such populations as polynesians and the Australian aboriginals.

The ***mechanism of race differentiation*** or divergence may be described in simple terms as follows. When a species covers a relatively large geographical area, it is likely to be divided into several geographically separated subunits or populations. ***Geographical separation*** is achieved when two populations of a species do not overlap, *i.e.,* their borders do not meet. Such a separation reduces the chances of gene-flow between the populations. Further, the environments to which the different geographically separated populations of a species are subjected are likely to differ. This will subject them to natural selection for adaptation to the prevailing environments, which is likely to alter their gene frequencies. Such alterations in gene frequencies of populations inhabiting different localities have been well documented in numerous instances. This explanation for race formation assumes that the differences in gene frequencies among races of a species have an adaptive value. These differences in gene frequencies may often be reflected in distinctive differences in the morphological features of different races, *e.g.,* in the case of human races, skin colour, eye colour, facial features, etc. Thus races are dynamic rather than static entities; they become more and more distinct during the course of evolution, and finally become established as separate species.

42.12. SPECIATION

A ***species*** consists of all such individuals, which intermate freely among themselves, but ordinarily do not interbreed with the members of other species. The process of formation of such groups possessing reproductive isolation is called ***speciation.*** Speciation begins with race formation, and *as a rule the races of a species are* ***allopatric,*** *that is, geographically isolated and occupying different territories.* But members of different species may be either allopatric or ***sympatric,*** that is, occupying the same territory or locality. Even when two species are sympatric, generally there is no intermating between them due to various reproductive isolation mechanisms. *Thus the crucial step, which converts two races of a single species into two distinct species is the development of reproductive isolation between them. Clearly, then, geographical isolation initiates speciation by initiating race formation but is culminated by the development of reproductive isolation.*

42.13. REPRODUCTIVE ISOLATION MECHANISMS

Any genetically determined feature of two or more populations, which restricts or prevents their intermating is a ***reproductive isolation mechanism.*** Meyer has classified these mechanisms into two broad categories: (1) premating, that is, those operating before the occurrence of fertilization, and (2) postmating, *i.e.,* those operating after fertilization has been effected.

42.13.1. Premating Isolation Mechanisms

The ***premating mechanisms*** may be due to one of the following: (1) seasonal or habitat differences, (2) behavioral differences, and (3) due to mechanical reasons. In

Mechanisms of Reproductive Isolation

A. *Premating Isolation*
- Seasonal isolation
- Habitat isolation
- Behavioral or sexual isolation
- Mechanical isolation

B. *Post-mating isolation*
- Gametic mortality
- Zygotic mortality or hybrid inviability
- Hybrid sterility

the case of ***seasonal*** or ***habitat isolation,*** potential mates from the two species do not meet because they flourish in different seasons or habitats. For example, *Tradescantia canaliculata* and *T. subaspera* flower in different seasons; in addition, one species is restricted to deep shade, while the other grows in sunlight. ***Behavioral*** or ***sexual isolation*** involves differences in the courtship patterns of different species. As a result, females of a species prefer males of the same species and interspecific matings are rather rare. Even morphologically similar animal species, *e.g., D. melanogaster* and *D. simulans* (often called *sibling species* due to their morphological similarity) may show extremely effective sexual isolation. In case of ***mechanical isolation,*** matings between individuals of the two species are attempted, but fertilization cannot be achieved because their genitalia do not fit together. For a long time, this type of isolation was considered to be the primary isolating mechanism in animals. However, there is little evidence that matings in which genitalia are markedly different are ever seriously attempted.

42.13.2. Postmating Isolation Mechanisms

The postmating isolation mechanisms may involve one of the following: (1) gametic mortality, (2) zygotic mortality or hybrid inviability, and (3) hybrid sterility. In case of ***gametic mortality,*** either the egg or the sperm is destroyed in interspecific crosses. For example, in plants, pollen grains may not germinate on foreign stigma or pollen tubes may fail to grow in foreign style. In many cases, however, fertilization takes place, but either the zygotes do not develop or the hybrid individuals have poor viability. This is known as ***zygotic mortality*** or ***hybrid inviability.*** There are numerous examples of this type in both animals and plants. For example, hybrid seeds obtained from the cross between *Linum perenne* and *L. austriacum* ordinarily fail to germinate unless rescued through embryo culture. In many species crosses, the hybrids are normal but sterile; this is termed as ***hybrid sterility.*** For example, the F_1 hybrid males from the cross between *Drosophila pseudoobscura* and *D. persimilis* are completely sterile.

Generally, more than one reproductive isolation mechanisms develop between species. For example, *D. pseudoobscura* and *D. persimilis* exhibit (1) habitat isolation, (2) sexual isolation as well as (3) hybrid male sterility. Such multiple isolation mechanisms reinforce each other and effectively prevent intermating between the concerned species.

42.13.3. Origin of Reproductive Isolation

Reproductive isolation mechanisms may originate either accidently or due to natural selection for these mechanisms because of weak and defective hybrids

between races. When two races have diverged sufficiently either due to geographical separation (***allopatric speciation***) or due to distruptive selection in the same locality (***sympatric speciation***), they may be expected to have accumulated different gene combinations making them highly adapted to their respective environments. When two such races become sympatric, *i.e.,* come to occupy the same territory, intermating between them would take place. However, the 'hybrids' between such races may be expected to be weaker or less adapted than the two parental races, and natural selection would act against the hybrids. Therefore, natural selection would favour those gene combinations that reduce intermating between the two races since this would increase the fitness of the two populations. Thus genes would be gradually accumulated, which would establish a reproductive isolation between the two races.

That this may be so can be exemplified by the following results. *D. pseudoobscura* and *D. persimilis* show sexual isolation at higher temperatures, but this isolation is quite weak at lower temperatures. An experimental population was constructed by bringing together equal numbers of flies (half male and half female) of the two species. This population was kept at low temperature; in each generation, the interspecific hybrid flies were identified and removed from the population. There was a gradual decrease with the generation in the number of hybrid flies produced in this population. Finally, the sexual isolation at low temperature became as strong as that at the higher temperature.

42.14. ORIGIN OF ADAPTATIONS

It is believed that adaptations originate in response to natural selection pressure on populations. When a population is subjected to a new selection pressure, it is expected that it will develop novel adaptations to the new environment. This has been demonstrated in laboratory cultures of microorganisms. Bacteria are commonly used for such studies because of their rapid generation turnover, limited space requirement, easy and effective selection techniques, etc. In various experiments, bacterial strains were grown on a new carbon source, *e.g.,* xylitol, or a new nitrogen source, *e.g.,* butyramide, which the bacterial cells are unable to metabolize properly. In many cases, the bacterial cells became increasingly efficient in the utilization of such unusual substrates. Often this was achieved by genetic changes due to which initially 'inefficient' enzymes acting on the new substrates (since these enzymes were used for some other purposes) became adapted to their utilization. The genetic changes responsible for such adaptations may be one of the following.

1. A regulatory mutation making the synthesis of the 'inefficient' enzyme constitutive so that it is produced in large amounts.
2. A regulatory mutation making this enzyme inducible by the new substrate.
3. Mutations in the structural gene producing this enzyme, which make the enzyme more 'efficient' in the utilization of new substrate.
4. Mutations that facilitate an unimpeded entry of the substrate into the bacterial cells.

Campbell and coworkers have shown that an enzyme generally not involved in the hydrolysis of lactose can become adapted to perform this task in *E. coli.* They used an *E. coli* strain in which the gene coding for β-galactosidase, the enzyme catalyzing hydrolysis of lactose, is deleted. Consequently, this strain is unable to

utilize lactose as a carbon source. When this Z^{del} (deleted *z* gene) strain was grown on a lactose containing medium for one month, it became capable of utilizing lactose, although inefficiently. Further growth and selection among these new lactose-utilizing cells yielded a more efficient lactose-utilizing strain, called *Ebg* (from *evolved* β-*galactosidase*). The lactose-hydrolysing enzyme of this *Ebg* strain is completely different from β-galactosidase; the new enzyme has a greater molecular weight, different immunological properties, and maps at a locus (at 59 min position) different from that for the β-galactosidase (at the 10 min position). These findings clearly demonstrate that an enzyme may take up a new function when subjected to appropriate selection pressure.

Often a new adaptation may arise in several different ways. For example, resistance in many insects to insecticides like DDT may arise due to one or more of the following reasons.

1. An increase in the lipid content of body enables the fat-soluble DDT to be separated from other parts of the organism.
2. Evolution of enzymes, which breakdown DDT to relatively less toxic products.
3. A reduction in toxic response of the nervous system to DDT.
4. Reduced permeability of the insect cuticle to DDT.
5. A changed behaviour, which reduces the chances of contact with DDT.

It is, therefore, not surprising that insecticide resistance is often due to several genes. In *Drosophila,* DDT resistance is polygenically determined, the genes for resistance being located in three of the four of its chromosomes.

QUESTIONS

1. Define Evolution. Briefly describe the essential features of various theories of evolution.
2. Explain the gradualistic and punctuational models of evolution and discuss their relative merits.
3. Describe the various mechanisms for the fixation of mutant alleles in populations with special reference to selectionist and neutralist concepts.
4. Discuss the various levels at which genetic variation may be assessed and describe the various sources of genetic variation.
5. Briefly describe the mechanisms of race and species formation.
6. Define reproductive isolation. Briefly describe the various types of reproductive isolation mechanisms and discuss a possible mode of origin of such mechanisms.
7. Define adaptation. How can molecular and other adaptations originate in living organisms.
8. Define natural selection. Describe the different types of selection and their effects on populations.
9. Define genetic polymorphism. Describe various types of polymorphism and the mechanisms that are responsible for them.
10. Write short notes on the following:
    (*i*) genetic polymorphism, (*ii*) natural selection,
    (*iii*) race formation, (*iv*) speciation,
    (*v*) genetic load, (*vi*) Darwinism,
    (*vii*) micro- and macro-evolution, (*viii*) reproductive isolation,
    (*ix*) adaptation.

43

Genetics and Man

43.1. INTRODUCTION

Man belongs to the genus *Homo,* which has a single species called *sapiens sapiens.* The genetic machinery of man and the mechanisms of repair, replication and distribution of the genetic material are comparable to any other plant or animal species. The single remarkable difference between man and the rest of the biological world appears to be the faculty of thinking and reasoning. Therefore, man can be viewed in two distinct picture frames: *firstly,* as an organism, and *secondly,* as the manipulator of genetic processes.

43.2. MAN THE ORGANISM

The species *H. sapiens sapiens* may be divided into several races, *e.g.,* mongoloid, negroid, etc., based on facial and other morphological features. Many investigators have claimed racial differences in *IQ* tests, which according to others may as well be due to ethnic, cultural and economic differences among the races studied.

43.2.1. Cytogenetics

The human chromosome number is $2n = 46$ (Fig. 43.1). These chromosomes

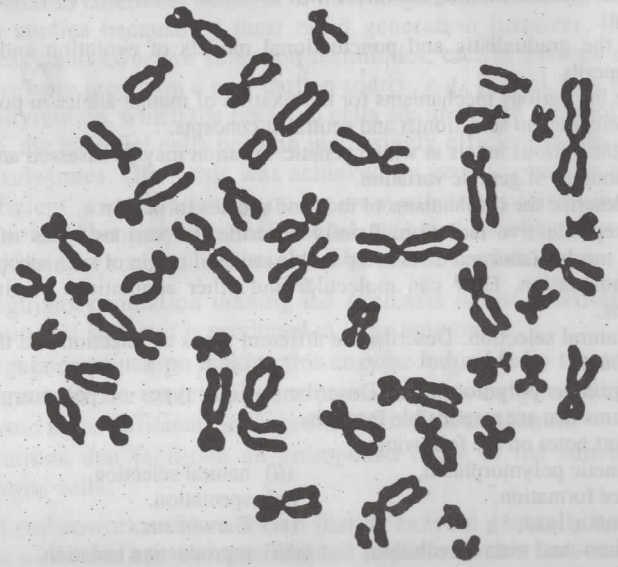


Fig. 43.1. Metaphase (mitotic) chromosomes of a normal human male (preparation from a dividing leucocyte). From J.L. Hammerton, *Int. Rev, Cytol* 12:1 (1961).

Human Cytogenetics

- $2n = 46$; chromosomes divided into 7 groups (A – G).
- Female XX, Male XY.
- Females show Barr body.
- Chromosome banding, etc. enable identification of every chromosome.
- *Numerical chromosonal aberrations* :
  - (*i*) Down's syndrome (trisome 21).
  - (*ii*) Patau syndrome (trisome 13).
  - (*iii*) Edward syndrome (trisome 18).
  - (*iv*) Metafemale (XXX).
  - (*v*) Klinefelter syndrome (XXY, XXXY etc.).
  - (*vi*) Turner syndrome (XO).
  - (*vii*) Triploid (3*x*).
- Only 10% of aneuploid zygotes survive.
- *Structural chromosonal aberrations* :
  - (*i*) translocations, deletions and inversions.
  - (*ii*) serious effects: leukaemia, cancer, etc.

are classified into seven (A-G) groups on the basis of their relative lengths and the centromeric positions (Fig. 43.2). Generally, human beings have 22 pairs of autosomes (designated as chromosomes 1-22) and one pair of sex chromosomes. Females have two *X* chromosomes, while males have one *X* and one *Y* chromosome. The *X* chromosome (denoted as *X*) is similar to group C chromosomes, and is included in that group. The *Y* chromosome (denoted by *Y*) is included in group G, being similar in morphology to the chromosomes of this group. Secondary constrictions are present in chromosomes 13, 14, 15, 21 and 22, the satellites forming dot-like extensions of the short-arm.

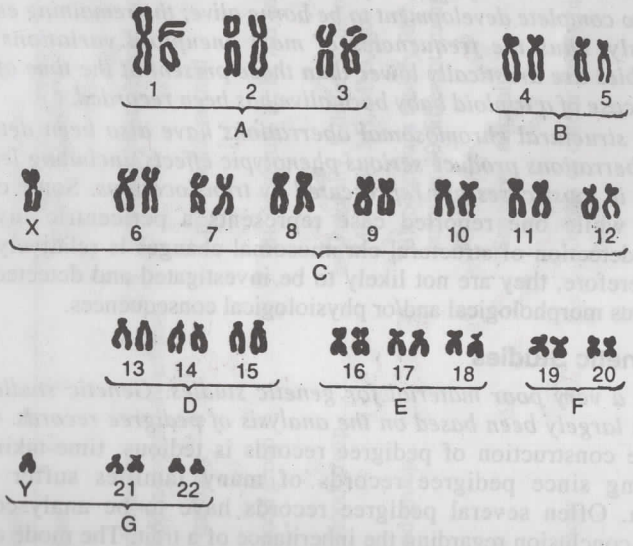


Fig. 43.2. Karyotype of a normal human male (based on the metaphase plate shown in Fig.43.1). From J.L. Hammerton, *Int. Rev. Cytol.* **12** :1 (1961).

43.2.1.1. Sex Chromatin

In females, one random *X* chromosome undergoes heterochromatinization quite early during embryo development. This serves as a means of dosage compensation. It also produces the ***Barr body*** or ***sex chromatin*** observable as a relatively large dark-staining regular-shaped body in somatic interphase nuclei of females.

43.2.1.2. Chromosome Banding

Distinct bands appear in metaphase chromosomes when they are processed according to specific staining procedures. The banding techniques have permitted an extraordinary resolution of human chromosomes. Each chromosome arm can now be divided into few to several distinct sections, which may be further subdivided into few to many subsections. Banding techniques coupled with uptake of radioactive thymidine (to identify late-replicating regions of chromosomes) permit a clear-cut identification of every single human chromosome.

43.2.1.3. Chromosomal Aberrations

Both numerical and structural chromosomal aberrations occur in variable frequencies in human populations. Numerical aberrations have been studied in a much greater detail. Trisomy for both smaller autosomes (Down's syndrome, trisome 21; Patau syndrome, trisome 13; Edward syndrome, trisome 18) as well as allosomes (*XXX*, triplo-*X* female or metafemale; *XXY, XXXY* and others, Klinefelter syndrome) occur in various frequencies, but monosomy is known for only sex chromosomes (*XO*, Turner syndrome; *XY*-the long arm of *Y* deleted, Turner syndrome; *YO*, does not survive). *Trisomy for the larger autosomes and monosomy for all the autosomes are so deleterious that embryos carrying these anomalies do not survive.*

Only a small fraction of the aneuploid zygotes (*about* 10% *in the case of trisome* 21) *survive to complete development to be borne alive; the remaining embryos abort spontaneously. Thus the frequencies of most aneuploid variations detected in newborn babies are drastically lower than those present at the time of conception. At least one case of a triploid baby born alive has been recorded.*

Several structural chromosomal aberrations have also been detected. These structural aberrations produce serious phenotypic effects, including leukaemia and cancer, and in most cases are represented by translocations. Some cases involve deficiency, while one reported case represents a pericentric inversion. The cytological detection of structural chromosomal changes is relatively much more tedious. Therefore, they are not likely to be investigated and detected unless they lead to serious morphological and/or physiological consequences.

43.2.2. Genetic Studies

Man is a very poor material for genetic studies. Genetic studies on human beings have largely been based on the analysis of pedigree records. Collection of data for the construction of pedigree records is tedious, time-taking and often disappointing since pedigree records of many families suffer from gappy information. Often several pedigree records have to be analysed to reach a dependable conclusion regarding the inheritance of a trait. The mode of inheritance of some human traits, both morphological and biochemical, is summarised in Table 43.1.

TABLE 43.1
Genetic control of some human traits

Character	*Alternate forms*	*Genetic control*	*Remarks*
Albimism	Albina, coloured	Albinism is monogenic autosomal recessive (*c*)	Inability to produce melanin.
Achondroplastic (chondrodystrophic) dwarfness	Achondroplastic dwarf, normal	Dwarfness due to a single dominant gene (*D*)	*D* allele is recessive lethal.
Crown baldness	Crown bald, normal	Monogenic, sex-influenced	Crown baldness dominant in males.
ABO blood group	Blood groups A, B, AB and O	Alleles I^A, $I^B$and *i* of a single gene	Codominance; encodes galactos-transferase enzyme.
MNSs blood group	Generally, 9 different blood groups MS, MSs, Ms, MNs, etc.	Four alleles of a single gene	codominance
Rh blood group	Blood groups mainly Rh^+ and Rh^-; but over 30 different groups known	Multiple alleles of a single gene	Responsible for haemolytic anaemia of newbornes
Favism	Sensitive, normal	Sex-linked recessive	Haemolytic anaemia in response to pollen and/or beans of *Vicia faba*
Skin colour	Grades of colour from white to black	Polygenic, positive alleles produce colour	A marked environmental effect; exposure to sunlight enhances colour

Genetic Studies

- Based on pedigree analyses.
- Humans are poor material for genetic studies.
- *Genome project* : Structural and functional map of human genome.
- Over 200 different genetic diseases.
- *Several blood groups* : ABO, Rh, MN, Duffy, P, Lewis, etc.
- *Autoimmune disease* : Antibodies produced against self-antigens, including DNA.
- *DNA fingerprinting* : Identification of victims, criminals, parents etc.
- *Histocompatibility* (*tissue graft rejection*) : Main HL-A system; also ABO and other blood groups.
- *Immune response* : Antibodies produced against foreign antigens.

A *human genome project* has been launched as a cooperative project involving several countries to prepare detailed structural and functional map of human genome using sophicated techniques of recombinant DNA technology. A draft sequence of the human genome was published in June, 2006. The conclusions from the draft sequence are summarised in chapter 26.

43.2.2.1. Genetic Diseases

Human species is affected by over 3,500 different diseases produced by mutant genes; such diseases are called ***genetic*** or ***hereditary diseases*** (Table 43.2). The frequency of occurrence of these diseases in the population may range from 4×10^{-4} (cystic fibrosis) through 10^{-4} (haemophilia, Huntington's chorea, albinism, phenylketonuria) to 5×10^{-6} (galactosemia) or even lower. Most of the genetic disorders are quite deleterious and, in many cases, death may occur before the reproductive age. But in some cases, *e.g.*, Huntington's chorea, the onset of disease may be well after the reproductive age permitting the affected persons to reproduce and transmit their genes to the next generation.

Genetic diseases not only claim lives of the afflicted individuals, they often cause tremendous suffering to such persons. These diseases also generate considerable emotional stress in otherwise normal relatives of the victims, may require large sums of money for management of victims, and often specialised institutions are required for their care. In most developing countries, adequate facilities for appropriate management of such victims may not exist.

43.2.2.2. Blood Groups

Human body responds to a variety of compounds, called ***antigens,*** by producing specific proteins (known as ***antibodies***) against them. Antibodies bind to the specific antigens in response to which they are produced (***antigen-antibody interaction***). Various kinds of antigens are present on the surfaces of blood cells and body tissue cells.

The ability of human body to produce antibodies against foreign antigens offers the only protection against viral diseases. It may be noted that human body does not ordinarily produce antibodies against its own antigens. But human body may rarely

TABLE 43.2
Some genetic diseases of man

Disease	*Symptoms*	*Loss in the activity of enzyme/Remarks*
Acatalasia	Oral gangrene	Catalase
Alkaptonuria	Darkening of urine, arthritis,	Homogentisic acid oxidase
Huntington's chorea	Progressive mental degeneration	Glutamic acid decarboxylase
Phenylketonuria	Mental retardation, excretion of phenylalanine in urine	Phenylalanine dehydrogenase
Galactosemia	Liver enlargement, galactose in urine and blood, cataract	Galactose-1-phosphate uridyl transferase
Cystic fibrosis	Excretion of highly viscous mucous; exocrine gland deficiencies	Beta glucourinidase
Sickle cell anaemia	RBC's become sickle-shaped under low O_2 concentration; anaemia and other symptoms	Amino acid replacement in haemoglobin β-chain
Colour blindness	Inability to distinguish between certain colours	Several types, genetic control rather complex, some forms, *e.g.*, protan defect (inability to perceive green colour), sex-linked.
Haemophilia A (sex-linked)	Lack of blood clotting due to which minor cuts bleed profusely	Reduction in the amount of anti-haemophilic factor
Myopia	Near-singhtedness	Sex-linked
Hunter syndrome	Mental retardation, abnormal hairiness, characteristic facial features including a large protruding tongue	Abnormal accumulation of mucopolysaccharides in cells; sex-linked recessive.

produce antibodies againsts its own antigens, thereby, producing ***autoimmune diseases***. There is some evidence that autoimmune diseases have a genetic basis.

Erythrocytes or RBCs (red blood corpuscles) contain on their surface several types of antigens. These antigens give rise to various blood group systems; some of the commonly detected systems are ABO, MNSs, Rh, P, Lewis, Duffy, etc. *Blood group antigens are expressed even before the child is borne and their expression is not affected by the environment. A majority of the blood group systems show multiple alleles.*

Blood groups are particularly important in blood transfusions in order to avoid antigen-antibody interaction between recipient and donor bloods. The consequences of incompatible blood transfusion can be quite serious. Ordinarily, ABO and Rh blood group compatibilies are carefully checked before a blood transfusion. Rh blood group is important from another point of view. Marriages between Rh^- females and Rh^+ males may produce babies suffering from haemolytic anaemia. Such incompatibilities may even lead to miscarriages. Generally, *the first pregnancy is not affected by Rh incompatibility; only the second and subsequent foetuses are exposed to the risk of haemolytic anaemia.*

43.2.2.3. Disputed Parentage, and Identity of Criminals and Victims

Disputes often arise as to the paternal parentage of a child. A comparison of blood groups of the two parents and the child may provide definite evidence of a lack of relationship between the child and the accused father. The data on blood groups may be combined with other sensitive and highly reliable and reproducible information, such as DNA fingerprinting. ***DNA fingerprinting*** is also used to identify victims of crimes/accidents and also the criminals. The approach, in simple terms, is briefly outlined below.

1. A sample of DNA from the accused/victim/suspect is obtained. Common sources of DNA include blood, semen, solid tissues, blood stains, semen stains, hair roots, etc.
2. The DNA is used to analyze certain highly variable and polymorphic regions of genomic DNA.
3. The analysis is based on RFLP (restriction fragment length polymorphism), PCR (polymerase chain reaction)-based markets, dot blot assay and, ultimately, base sequence analysis.
4. These data are analyzed and compared to arrive at a suitable decision. It is extremely unlikely that the DNA fingerprint (or profile) of any two individuals will be identical. The technique of DNA fingerprinting is extremely efficient, extraordinarily precise and requires very small amounts of DNA, which can be obtained even from old blood stains, bones, etc.

43.2.2.4. Histocompatibility

Transplants of skin and other tissues from one individual into another are rejected, due to the production of antibodies in the host against the antigens present in the transplanted tissues; such a reaction is known as ***histocompatibility reaction***. In general, subsequent tissue explants from the same donor are rejected more rapidly than the first transplant. Tissue transplants are generally successful only between identical twins.

In human beings, histocompatibility is mainly determined by *HL-A* system. In addition, ABO and other blood group systems are also involved. The rejection of tissue transplants is prevented by administering into the host (the recipient of transplant) system 'immunosuppressive' drugs. These drugs prevent the growth and division of antibody producing cells. As a result, the patient becomes highly susceptible to bacterial and viral infections and to cancer.

43.2.2.5. Immune Response

When human body is exposed to an antigen found on the cell surface of a pathogen, it produces antibodies specific to this antigen. In addition, a population of 'memory' B cells are produced; these cells remain in the body for a long time, and begin to produce the antibody soon after they are exposed to the same antigen again. Therefore, when such an individual is exposed to the same pathogen again, the antibodies specific to this pathogen are produced by the 'memory' cells. These antibodies would inactivate the pathogen through antigen-antibody interaction. In this way, the individual would be protected against the concerned disease. This ability of human body to acquire immunity against pathogens through a prior exposure to antigens specific to the concerned pathogen is known as ***immune response***. This forms the basis for immunisation of human beings against various dreaded diseases, *e.g.*, small pox, polio, cholera tetanus, diphtheria, measles, etc. *The vaccines used for immunisation against various diseases generally contain some antigens specific to the concerned pathogens.*

Tumour

- Produced by cells capable of uncontrolled growth.
- *Types* : (1) benign and (2) malignant.
- *Carcinogens*: Agents that induce tumour/promote tumour growth.
- Changes leading to cancer formation:
  1. Immortalization.
  2. Transformation.
  3. Metastasis.
- Typically, 6-7 events are needed to produce cancer.

43.3. GENETICS OF CANCER

Cancer is produced by cells that have acquired the ability for uncontrolled growth and to grow in inappropriate locations; such cells are said to be ***cancerous*** or ***tumorigenic***. Uncontrolled growth in noncirculatory tissues produces solid tumours. When a tumour becomes cancerous or malignant, its cells detach and migrate to other parts of body where they produce ***secondary tumours***; this phenomenon is called ***metastasis***. ***Benign*** or ***noncancerous tumours*** do not show metastasis. Usually, multiple events (typically, 6–7) are required to produce cancer; these events may be triggered by genetic or environmental changes. A variety of agents, called ***carcinogens***, either increase the frequency of tumour formation or promote tumour growth. Rarely, cancer may be caused by a single genetic change, but in some cases multiple genetic changes have been identified.

43.3.1. Type of Cellular Changes

The changes that ultimately lead to tumour formation are grouped into the following 3 types: (1) immortalization, (2) transformation and (3) metastasis.

Immortalization

- Cells cultured *in vitro* become capable of indefinite growth.
- *Features of immortalized cells*:
  1. Require a solid surface to attach to for growth.
  2. Need serum.
  3. Show contact inhibition.
  4. Are flat and extended in shape.
- As a result, immortalized cells grow as monolayers.

43.3.1.1. Immortalization

When normal vertebrate cells are cultured *in vitro*, they grow for ~50 cell generations, before they enter a period of crisis. During ***crisis***, most of the cells die. However, some cells survive crisis, and have the ability of indefinite growth; this acquisition of the property of indefinite growth is called ***immortalization***. Immortalized cells constitute ***permanent*** or ***established cell lines,*** which have the following main features : (1) anchorage-dependence, (2) need for serum, (3) growth upto a limited cell density, (4) contact inhibition and (5) flat and extended shape. As a result, they grow only as a single cell thick layer, called ***monolayer***, that is adhered to the surface of culture vessel (called ***substrate***). These cells do not form tumours when injected into test mice.

Transformation

- *Cultured vertebrate cells become*:
  1. Anchorage-independent.
  2. Serum-independent.
  3. Free from contact inhibition.
  4. Round in shape.
  5. Produce tumours *in vivo*.
  6. Form foci *in vitro*.
- *Caused by*:
  1. Viruses.
  2. Cellular oncogene activation.
  3. Mutant suppressor genes.

43.3.1.2. Transformation

Transformation refers to cultured vertebrate cells becoming anchorage and serun-independent, and free from contact inhibition. Transformed cells become round in shape, and form a thick mass of cells in culture called ***focus***. When injected into appropriate test animals, they may form solid tumours. Transformation may be induced by activation of a cellular proto-oncogene, introduction of a viral oncogene by a retrovirus, or by loss of function of tumour suppressor genes.

43.3.1.3. Metastasis

It is the ability of a tumour cell to migrate from the tumour and establish a colony in another part of the body. In cultured cells, the cell migrates from the colony and forms a new colony at another location. Molecular basis of metastasis is not well known. Metastasis is the most damaging to the organisms and is exhibited by malignant tumours.

43.3.2. Transformation by Viruses

Infection by many DNA and RNA viruses may lead to tumour formation; these viruses are called ***tumour viruses*** or ***transforming viruses***. Tumour viruses are of the following two types : (1) DNA tumour viruses and (2) tumour retroviruses.

43.3.2.1. DNA Tumour Virus

Those viruses that contain DNA genomes and cause tumour formation following infection are called ***DNA tumour viruses***. Examples of DNA tumour viruses are, polyomaviruse, human papillomaviruses (HPV), adenoviruses and Epstein-Barr virus. *The ability to cause transformation is due to a single or a group of proteins produced early during the lytic cycle of viruses*. Transformation is caused by integration of the whole viral genomes, or of those parts of viral genomes that contain the early genes, into the host cell genomes (Fig. 43.3). These viruses cause transformation of only those cells that can not be productively infected by the virus. In such cells, virus replication is incomplete and host cell lysis does not take place.

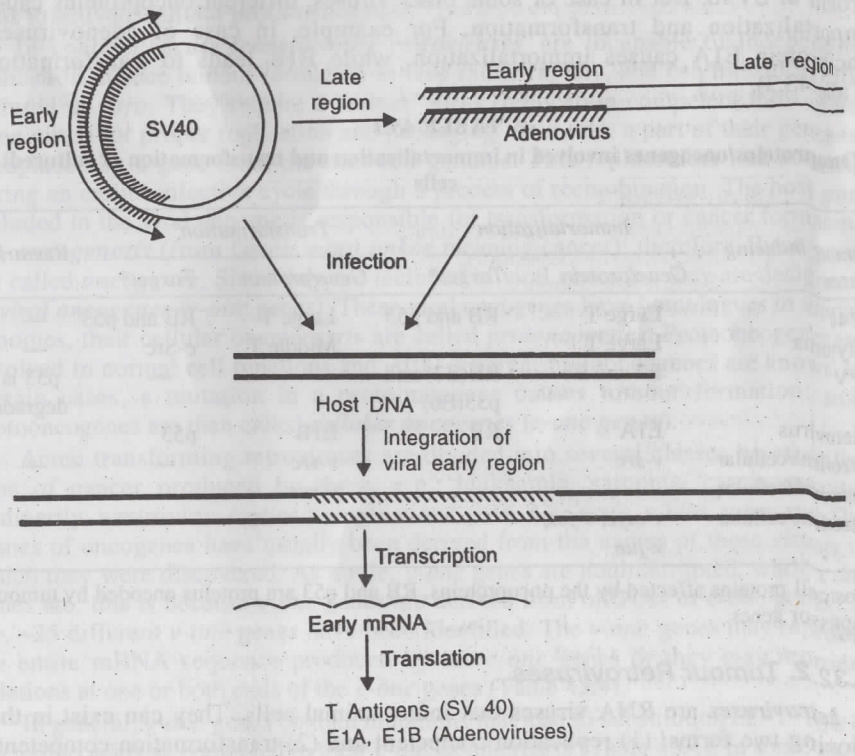


Fig. 43.3. A model for transformation by DNA tumour viruses. At least the early region of viral genome is integrated and the early protein products interact with cellular proteins to cause transformation.

43.3.1. Type of Cellular Changes

The changes that ultimately lead to tumour formation are grouped into the following 3 types: (1) immortalization, (2) transformation and (3) metastasis.

Immortalization

- Cells cultured *in vitro* become capable of indefinite growth.
- *Features of immortalized cells*:
  1. Require a solid surface to attach to for growth.
  2. Need serum.
  3. Show contact inhibition.
  4. Are flat and extended in shape.
- As a result, immortalized cells grow as monolayers.

43.3.1.1. Immortalization

When normal vertebrate cells are cultured *in vitro*, they grow for ~50 cell generations, before they enter a period of crisis. During ***crisis***, most of the cells die. However, some cells survive crisis, and have the ability of indefinite growth; this acquisition of the property of indefinite growth is called ***immortalization***. Immortalized cells constitute ***permanent*** or ***established cell lines,*** which have the following main features : (1) anchorage-dependence, (2) need for serum, (3) growth upto a limited cell density, (4) contact inhibition and (5) flat and extended shape. As a result, they grow only as a single cell thick layer, called ***monolayer***, that is adhered to the surface of culture vessel (called ***substrate***). These cells do not form tumours when injected into test mice.

Transformation

- *Cultured vertebrate cells become*:
  1. Anchorage-independent.
  2. Serum-independent.
  3. Free from contact inhibition.
  4. Round in shape.
  5. Produce tumours *in vivo*.
  6. Form foci *in vitro*.
- *Caused by*:
  1. Viruses.
  2. Cellular oncogene activation.
  3. Mutant suppressor genes.

43.3.1.2. Transformation

Transformation refers to cultured vertebrate cells becoming anchorage and serun-independent, and free from contact inhibition. Transformed cells become round in shape, and form a thick mass of cells in culture called ***focus***. When injected into appropriate test animals, they may form solid tumours. Transformation may be induced by activation of a cellular proto-oncogene, introduction of a viral oncogene by a retrovirus, or by loss of function of tumour suppressor genes.

43.3.1.3. Metastasis

It is the ability of a tumour cell to migrate from the tumour and establish a colony in another part of the body. In cultured cells, the cell migrates from the colony and forms a new colony at another location. Molecular basis of metastasis is

not well known. Metastasis is the most damaging to the organisms and is exhibited by malignant tumours.

43.3.2. Transformation by Viruses

Infection by many DNA and RNA viruses may lead to tumour formation; these viruses are called ***tumour viruses*** or ***transforming viruses***. Tumour viruses are of the following two types : (1) DNA tumour viruses and (2) tumour retroviruses.

43.3.2.1. DNA Tumour Virus

Those viruses that contain DNA genomes and cause tumour formation following infection are called ***DNA tumour viruses***. Examples of DNA tumour viruses are, polyomaviruse, human papillomaviruses (HPV), adenoviruses and Epstein-Barr virus. *The ability to cause transformation is due to a single or a group of proteins produced early during the lytic cycle of viruses*. Transformation is caused by integration of the whole viral genomes, or of those parts of viral genomes that contain the early genes, into the host cell genomes (Fig. 43.3). These viruses cause transformation of only those cells that can not be productively infected by the virus. In such cells, virus replication is incomplete and host cell lysis does not take place.

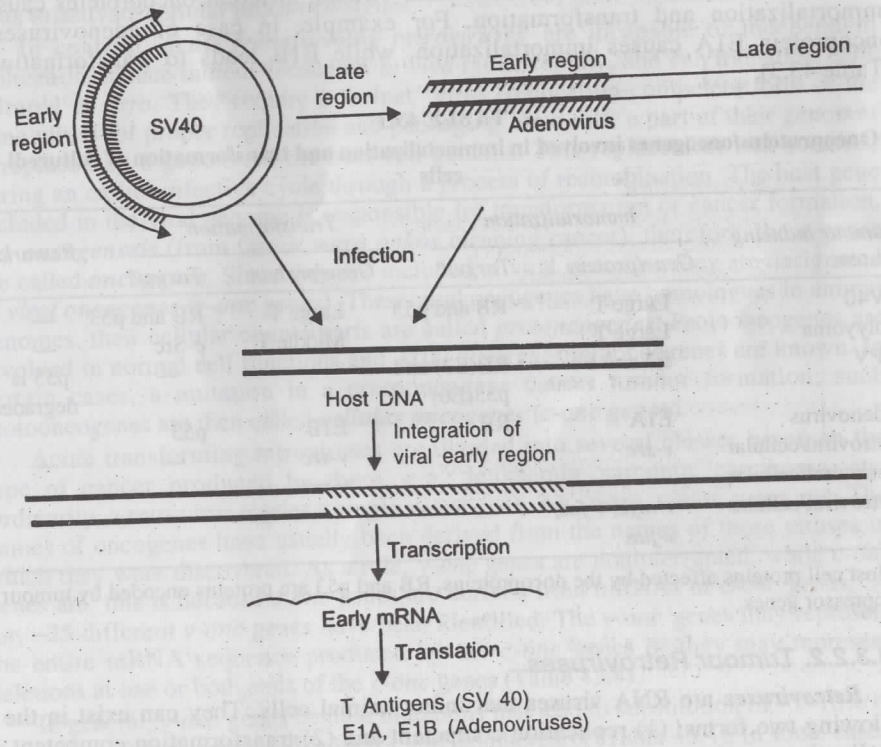


Fig. 43.3. A model for transformation by DNA tumour viruses. At least the early region of viral genome is integrated and the early protein products interact with cellular proteins to cause transformation.

Tumour viruses

- Viral infection produces tumour.
- *Types*:
  1. DNA tumour viruses.
  2. Tumour retroviruses.
- *DNA tumour viruses*:
  1. Transformation caused by integration into host genome of early region (at least) of viral genome.
  2. The viral proteins interact with cellular proteins.
  3. *Examples* : Polyomaviruses, HPV, adenoviruses, Epstein Barr virus.
  4. Transformation of only those cells that virus is unable to lyse.

The proteins encoded by the early regions of DNA tumour viruses interact with host cell proteins that ultimately causes quiescent (nondividing) cells to divide; these viral proteins are called ***oncoproteins***. In case of some viruses, the same oncoprotein causes both immortalization and transformation, *e.g.*, large-T (tumour antigen) protein of SV40. But in case of some other viruses, different oncoproteins cause immortalization and transformation. For example, in case of adenoviruses, oncoprotein E1A causes immortalization, while E1B leads to transformation (Table 43.3).

TABLE 43.3
Oncoproteins/oncogenes involved in immortalization and transformation of culturedl cells

Tumour inducing system	*Immortalization*		*Transformation*		*Remarks*
	Gene/protein	*Target**	*Gene/protein*	*Target**	
SV40	Large T	RB and p53	Large T	RB and p53	—
Polyyoma	Large T		Middle T	*c*-Src	—
HPV	E7 and E6	RB(E7) and p53(E6)	—	—	p53 is degraded
Adenovirus	E1A	RB	E1B	p53	
Retroviral/cellular	*v-src*	—	*v-src*	—	—
Retroviral/cellular	—	—	*v-ras*	—	—
Retroviral/cellular	*v-myc, v-fos, v-jun*	—	—	—	—

*Host cell proteins affected by the oncoproteins. RB and p53 are proteins encoded by tumour suppressor genes.

43.3.2.2. Tumour Retroviruses

Retroviruses are RNA viruses that infect animal cells. They can exist in the following two forms: (1) replication competent and (2) transformation competent. ***Replication competent*** form is able to replicate and multiply in the host cells, but *it does not cause transformation*. But ***transformation competent*** form causes transformation of host cells; this form, however, may or may not be able to replicate independently, *i.e.*, when present alone in a host cell. Retroviruses multiply by

Tumour Retroviruses

- RNA viruses that infect animal cells and cause tumour.
- *Forms*: 1. replication competent and
  2. transformation competent.
- *Replication competent form*:
  1. Normal retroviral life cycle.
  2. Long latent periods.
  3. Oncogenicity due to mutations that activate *c-onc* genes.
  4. Usually, associated with leukaemias.

integrating into the host cell genome. When they become integrated in the genome of germline cells, they are passed on to the progeny in the same way as any other gene of the host genome.

Replication competent tumour retroviruses follow normal retroviral life cycle. Their ability to cause tumours arises from individual mutations and other genetic changes in their genomes. The infectious particles produced by these viruses have long latent periods. They are usually associated with the induction of leukaemias, *e.g.*, feline leukaemia virus (FeLV). Tumour formation depends on the ability of the virus to activate cellular proto-oncogenes.

In contrast, ***acute transforming retroviruses*** are incapable of independent replication, induce tumour formation *in vivo* rather rapidly, and can transform cells cultured *in vitro*. They require a 'helper' virus (replication-competent form of the same virus) for proper replication and packaging. As a rule, a part of their genomes is replaced by a gene from the host cell genome. This replacement had occurred during an earlier infective cycle through a process of recombination. The host gene included in the viral genome is responsible for transformation or cancer formation, *i.e.*, ***oncogenesis*** (from Greek word *onkos* meaning cancer); therefore, these genes are called ***oncogenes***. Since they are included in viral genomes, they are designated as ***viral oncogenes*** (***v-onc*** genes). These viral oncogenes have homologues in animal genomes, their cellular counterparts are called ***protoncogenes***. Protooncogenes are involved in normal cell functions and ~100 different protooncogenes are known. In certain cases, a mutation in a protooncogene causes tumour formation; such protooncogenes are then called ***cellular oncogenes*** (***c-onc genes***).

Acute transforming retroviruses are divided into several classes based on the type of cancer produced by them, *e.g.*, leukaemia, sarcoma, carcinoma, etc. Ordinarily, a retrovirus carries a single *v-onc* gene, *e.g., v-src, v-mos, v-ras,* etc. The names of oncogenes have usually been derived from the names of those viruses in which they were discovered. As a rule, *v-onc* genes are noninterrupted, while *c-onc* genes are; this is because *v-onc* genes are derived from mRNAs of c-onc genes. So far, ~25 different *v-onc* genes have been identified. The *v-onc* genes may represent the entire mRNA sequence produced by the *c-onc* genes or they may represent deletions at one or both ends of the *c-onc* genes (Table 43.4).

In general, *v-onc* genes contain mutations that lead to substitution of ~1–10% of the amino acids as compared to their *c-onc* counterparts (Table 43.4). In some cases, these mutations are directly responsible for oncogenesis. But in some other cases, overexpression of the *v-one* genes causes tumour formation. In general, the level of expression of *v-onc* genes is typically much greater than that of the corresponding *c-onc* gene.

Transformation Competent Retroviruses

- Replication defective.
- Induce tumour *in vivo*.
- Cause transformation *in vitro*.
- A part of their genome is replaced by a host gene, called *v-onc* gene.
- *v-onc* genes have homologues in host cells, which are called *c-onc* genes.
- *v-onc* gene are:
  1. noninterrupted.
  2. complete or truncated mRNA sequence of *c-onc* genes.
  3. contain mutations.
- Level of expression much higher than that of *c-onc* genes.
- ~ 25 different *v-onc* genes known.

TABLE 43.4

The degree of relationship between some *v-onc* genes and their cellular counterparts (c-one genes)

Retrovirus	*v-onc gene*	*c-onc gene*	*Deleted region*	*Number of changed amino acids*
Moloney murine sarcoma (mouse)+	v-*mos*	*c-mos*	None	11
Harvey murine sarcoma (rat)	*H-ras*	*c-H-ras*	None	3
Feline sarcoma (PI-FeSV) (cat)	v-*sis*	*c-sis*	None	18
Avian myelocytomatosis (chicken)	v-*myc*	*c-myc*	None	2
Rous sarcoma (chicken)	v-*src*	*c-src*	*19 amino acids	16
Feline sarcoma (Sm-FeSV) (cat)	v-*fms*	*c-fms*	*50 amino acids	20
Avian erythroblastosis (chicken)	v-*erbA*	*c-erbA*	**12 amino acids	22

\+ The animals listed within parenthesis serve as hosts of the respective retroviruses.
* Deletion of C-terminal amino acids.
** Deletion of N-terminal amino acids.

43.3.3. Cellular Oncogenes

Cellular oncogenes (*c-onc*) have been identified by the following two approaches: (1) detection of homology with *v-onc* genes of sequences in the host cell genome, and (2) looking directly for cancer causing genes in the genomes of cancer cells by using their DNA in suitable transfection experiments.

Protooncogenes may become activated to cause tumours, *i.e.* function as cellular oncogenes. The oncogenic effect may be produced due to the following.

1. Mutation leads to a change in the activity of protooncogene product. Some oncogenes are activated by point mutations, *e.g.*, *c-ras*. In case of *c-ras*, amino acid substitution at position 12 or 61 makes it oncogenic.
2. Increased expression of protooncogene may itself be oncognenic. Increased expression may be the result of mutation (*e.g.*, in *c-ras*), amplification (*e.g.*, *c-myc, c-abl, c-myb,* etc.), insertion of replication competent setroviruses (*e.g., c-myc, c-myb, c-mos*, etc.), and translocation (*e.g., c-myc* in the case of Burkitt lymphomas in man).

3. In case of some protooncogenes, any one of the above two changes or both the changes may be oncogenic. For example, some mutant *c-ras* genes also contain a mutation in an intron that increases *c-ras* mRNA processing by 10-fold leading to increased level of *c-ras* expression.

Cellular Oncogenes

- Proto-oncogenes are involved in normal cell division or growth.
- Activated to become tumourigenic by:
  1. Mutation (changes activity of the protein).
  2. Increased expression due to mutation, amplification, retroviral insertion or translocation.
  3. Both of the above changes.

43.3.4. Function of Protooncogenes

The proteins encoded by cellular oncogenes can be grouped into the following seven categories on the basis of the functions performed by them (Table 43.5).

TABLE 43.5

Functions of different proto-oncogene-encoded proteins

Type of function	*Oncoprotein-specific function*		*Location*
Growth factors	c-Sis	PDGF B chain	Secreted proteins
	KS/HST	Related to EGF	
	Wnt1	Related to wingless	
	Int2	Related to FGF	
Growth factor receptors	c-ErbB	EGF receptor kinase	Transmembrane
	erbB23	EGF-like receptor kinase	Transmembrane
	c-Fms	CSF-I receptor kinase	Transmembrane
	c-Kit	Steel receptor kinase	Transmembrane
	Mas	Angiotensin receptor	Transmembrane
G protein/signal transduction	c-Ras	GTP-binding protein	Membrane associated
	Gsp/Gip	Gαs and Gαi	Membrane associated
Intracellular tyrosine kinase	c-Src	Tyrosine kinase	Membrane associated
	c-Abl	Tyrosine kinase	Cytosolic
	c-Fps	Tyrosine kinase	Cytosolic
Serine/threonine kinase	c-Raf	Ser/Thr kinase	Cytosolic
	c-Mos	Ser/Thr kinase	Cytosolic
Signalling	Crk	SH2/SH3 regulator	Cytosolic
	Vav	SH2 regulator	cytosolic
Transcription factors	c-Myc	HLH protein	
	c-Myb	Transcription factor	
	c-Fos	Leucine zipper protein	
	c-Jun	Leucine zipper protein	
	c-Rel	NF-κB family	
	c-ErbA	Thyroid hormone receptor	

1. Several of them are ***growth factors;*** such a protein is secreted by one cell and acts on other cells. In these cases, the mechanism of oncogenesis is not known.
2. Many protooncogenes produce ***transmembrance growth factor receptors*** that are protein kinases. Oncogenicity may result from constitutive activation of the kinase activity.
3. Several proteins form ***components of signal transduction*** pathways.
4. Proteins like c-Src, c-Abl, etc. serve as ***intracellular protein tyrosine kinases.***
5. Several proteins like c-Raf and c-Mos are ***cytosolic serine/threonine kinases.*** Oncogenic mutations probably increase or constitutively activate the kinase activities.
6. Proteins like Crk and Vav are involved in the ***later stages of signal transduction.***
7. Many protooncogenes encode ***transcription factors***, *e.g., c-myc, c-myb, c-fos,* etc.

Tumour Suppressor Genes

- Suppress cell division.
- Deletion/inactivation of both alleles leads to tumour formation.
- Examples : p^{53} and *RB* loci.
- RB protein binds to proteins needed for cell division; active protein inhibits cell division.
- Loss or inactivation of RB function relieves inhibition of cell division.

43.3.5. Tumour Suppressor Genes

Tumour suppressor genes suppress cell division and growth. Therefore, deletion or inactivation of both the alleles of such a gene is tumorigenic, *i.e.,* results in tumour formation. The two best characterized genes encode the proteins RB and p^{53}. A loss or inactivation **RB locus** is involved in retinoblastoma, a childhood disease in which tumour forms at retina, and some other forms of cancer. Protein RB is a nuclear protein, which is specifically phosphorylated at the end of $G1$ phase. It is dephosphorylated during mitosis, and remains in this state till the end of $G1$, when it is phosphorylated again. According to one model, unphosphorylated RB binds to cellular proteins that are needed for cell division; therefore active RB inhibits cell division. But phosphorylated RB is unable to bind these proteins; consequently, RB phosphorylation relieves the inhibition of cell division. Therefore, inactivation of both the alleles of RB allows the cells to enter division.

Gene p^{53} encodes a nuclear phosphoprotein, which most likely is an autotetramer. Protein p^{53} seems to be involved in a common control of cell proliferation. This protein binds DNA and activates transcription of some genes but suppresses that of some others; these genes possibly regulate cell cycle. Wild type p^{53} can suppress or inhibit the transformation of cells in culture by various oncogenes. Mutant forms of p^{53} fail to bind DNA; it is not known if this failure is related to the loss in tumour suppressor function.

Man the Manipulator

- Man changes the genetic makeup of other species (animal and crop improvement).
- Changes its own phenotype (euphenics) and genotype (eugenics).
- *Positive eugenics* : Controlled marriages not feasible; eutelegenics being used on a scale (USA).
- *Negative engenics* : Reduction in the frequency of harmful alleles.
  - (*i*) *Genetic counselling* : Advice to couples regarding having a child; antenatal diagnosis and abortion (if foetus has genetic disease).
  - (*ii*) *Genetic surgery* : Use of cloned genes to correct genetic defects: used in patients for disease cure; not used with human embryo.

43.4. MAN THE MANIPULATOR

In the course of natural evolution, the criterion of success and survival of various organisms was primarily their ability to adapt to the prevailing environment. This was the case until evolution produced man. Man has been making every effort since times immemorial to assume the charge of evolutionary processes of at least some species, including himself. He has made slow but steady progress in this direction, and his efforts have produced some remarkable results.

43.4.1. Animal and Crop Improvement

One of the, perhaps the most, important decisions of the early man, which later shaped his destiny was to tame certain plants and animals to his best advantage (***domestication***). Domestication of these species provided man with the material on which he could test and sharpen his tools for genetic modification of these species in order to increase their usefulness to himself. In these efforts, the objectives of man often ran counter to the dictates of natural selection operating on wild forms. As a consequence, the domesticated forms have diverged considerably from their wild relatives, and in most cases they are most likely to fare very poorly if allowed to return to the wild state and be subjected to natural selection.

Man has accumulated a vast body of scientific knowledge on animal and plant genetics and has devised ingenious breeding schemes for the improvement of these organisms. The successes in these breeding efforts have been phenomenal wherever a concerted and well-planned effort has been made. Hybrid varieties of crops and hybrid animal strains provide excellent examples of these successes. Man has even synthesized some successful promising crop species, *e.g., Triticale, Raphanobrassica,* etc. He has also domesticated many microorganisms for harvesting specific products, *e.g.,* antibiotics, etc. But these successes only activated the imagination of man by giving him the glimpses of the possible improvements through genetic manipulations in the domesticated organisms. In order to fulfill these expectations, man has been constantly busy devising more and more ingenious, sophisticated and powerful techniques, generally represented by the term ***biotechnology,*** some of which are briefly described in the next few chapters.

43.4.2. Euphenics

Man has been constantly trying to improve his own lot by adopting measures that either improve his phenotype or, in more recent years, his genotype. The various organised and individual educational and training efforts are aimed at improving the various skills of man making such trained persons enormously more competent than the untrained ones. The benefits of such trainings of both body and mind were quite obvious very early in the history of human civilization, and they have contributed immensely to the development of human culture and civilization. The development of very powerful scientific and technological knowledge is the outcome of these trainings and surely it does not reflect a concomitant alteration in the genotypes of modern man as compared to those of the earlier individuals.

These acquired skills and techniques have given man enough power to alter his own environment, often unfortunately (*e.g.,* pollution due to smoke, chemicals, effluents, etc.; working conditions and personal relationships, which induce/promote mental tension leading to peptic uncers and blood pressure, etc.). These changes, particularly during the late nineteenth and the twentieth centuries, have been rather rapid, and the pace of these changes is increasing markedly. However, the genetic makeup of man has not changed with the same pace so that he finds himself inadequate to cope up with his altered environment. An outstanding example is the human birth rate. Before the development of extensive medical facilities, production of more children by a couple was to the advantage of the human species in view of high infant mortality and low average life expectancy. But in most countries, the current infant mortality rates (*e.g.,* in Chile 10%, in USA 3%, etc.) are drastically lower than what they were before, and the average life span has considerably increased. However, the human reproductive capability has not showed a proportionate decline, which has led to the problems of ***population explosion.*** In order to check the global human population at a reasonable level, men and women have to be educated and trained to adopt family planning measures and exercise self-restraint in the matter of procreation.

The above discussions may be regarded as a part of euphenics since these measures aim at improving the human phenotype without altering his genotype. But in its original sense, as proposed by Lederberg, ***euphenics*** consists of nutritional and other measures, including surgery, that permit a normal development, function and, more importantly, reproduction by otherwise inferior or deleterious (even lethal) genotypes. Some good examples of euphenic measures are: (*i*) regular administration of insulin to diabetics, (*ii*) dietary regulation of alkaptonurics, (*iii*) rescue of haemophiliacs through regular blood transfusions, etc.

Euphenic measures allow reproduction by many genotypes that would have not done so without them. Many scientists, notably Muller, have expressed the fear that these measures may lead to an increase in the frequency of deleterious alleles in the population. In United States, at least 20-25% of newborns have either physical or mental defects, which account for about 15% of all infant deaths. It has been estimated that about 50% of all childhood deaths are the consequences of congenital (*i.e.,* from birth) defects or the complications arising from them. It is argued, and with sufficient reason, that a large proportion of these birth defects has a genetic basis. It may be pointed out that many genetic defects, *e.g.,* diabetes, muscular dystrophy, Huntington's chorea, etc., are expressed only later in life, and they are not included in the above estimates.

Thus the frequency of deleterious genes appears to be quite high in human populations. According to Muller, Crow and others, every normal individual in USA is expected to carry, in heterozygous state, deleterious genes, which are equivalent to 1-8 lethal genes. Muller expressed the fear that euphenic measures would lead to an increase in the frequency of such alleles to such an extent that the normal individuals of human race may have to direct all their efforts at the maintenance of their genetically defective kith and kin. However, there is little evidence so far to suggest that this fear may come true in the near future.

43.4.3. Eugenics

Even though euphenic measures may not convert a section of human population into genetic cripples, it is desirable to reduce the genetic load (***negative eugenics***) and to improve the genetic make up (***positive eugenics***) of human populations. Several wide ranging measures have been suggested in order to achieve one or both of these ends; some of these are briefly described below.

43.4.3.1. Positive Eugenics

Measures aimed at improving the genotypic constitution of human populations so that the individuals become better adapted and more fit than their predecessors, constitute ***positive eugenics.*** The phenomenal success of man in modifying the domesticated species provides sufficient basis for advocating similar programmes to be undertaken with the *Homo sapiens sapiens.* This would involve marriages between individuals possessing 'superior' phenotypes, and banning reproduction by the rest of the population (***controlled mating and reproduction***). The idea of controlled breeding was first suggested by Plato in his 'The Republic'. Later laws were inacted in the Nazi Germany that banned marriages between members of the 'superior' (Aryan) and 'inferior' races. The caste system in India seems to be based on similar considerations and marriages between members of different castes are not looked upon favourably.

It has been advocated that selective breeding may be able to improve characters like intelligence, aesthetic sensitivity, resistance to diseases and stresses, etc. Sufficient variability seems to exist for these traits, and they seem to have some genetic basis. However, (*i*) measurement of some of these traits is quite problematic (*e.g.,* aesthetic sensitivity), (*ii*) genetic and environmental effects on the expression of these traits are difficult to separate, and, above all, (*iii*) controlled mating and reproduction does not seem to be feasible under the present sociopolitical environment.

Another approach for positive eugenics is ***eutelegenics*** or ***germinal choice,*** which may be more acceptable than selective breeding. In this approach, semen samples from individuals of outstanding worth are stored in liquid nitrogen in ***sperm banks,*** and is made available to women for conception through artificial insemination (AI), preferably after the semen donors have passed away. AI seems to have found some social acceptance in USA for countering male infertility among married couples; more than 10,000 children have already been born in this way. It is hoped that children produced through eutelegenics would be superior at least for the traits in which the selected sperm donor was outstanding. But it may be pointed out that this scheme is similar to that for simple mass selection in maize (except that here the selection is based on the male parent, while in maize it is based on the

female parent). Therefore, it may be quite effective for relatively simply inherited traits, while it may have little or no effect on complex traits. Some of the genetic problems of such a scheme are: (1) measurement of the desirable trait, (2) genetic component (heritability) of the trait, and (3) the effects of segregation and recombination.

An approach that preserves the genotype of a superior individual is to place his diploid nuclei in enucleated eggs, and then to induce these eggs to follow the developmental pattern of normal zygotes. In this way, clones of superior individuals can be produced. An alternative approach is to induce females to lay unreduced (diploid) eggs (produced without meiosis) and to induce these eggs to develop into embryos and infants parthenogenetically (without fertilization). This approach would also give rise to clones of selected individuals but would be confined to females. However, techniques for implementing these proposals are not yet available, and there are several associated moral, social and other issues, which need to be resolved.

43.4.3.2. Negative Eugenics

Several measures have been contemplated to reduce the frequency of deleterious genes in human populations; such approaches are often referred to as ***negative eugenics.*** These measures represent the following two broad approaches: (*i*) reducing the reproductive success of individuals carrying deleterious genes without adversely affecting the quality of their lives, and (*ii*) to correct the genetic defects of such individuals through genetic surgery.

43.4.3.3. Genetic Counselling

Genetic counselling consists of educating prospective parents either suffering from or suspected to be heterozygous for some genetic disease on the risk of their children suffering from the same disease. It is hoped that such parents would voluntarily abstain from producing children; this may be encouraged by creating a suitable social environment. It is relatively easy for a trained clinician to identify people suffering from a genetic disease. But the identification of carriers (heterozygotes) for genetic diseases is much more difficult and, in many cases, not possible. Heterozygotes for genetic diseases are readily identifiable when they have either a reduced level of the concerned enzyme (*e.g.*, HGPRT enzyme in Lesch-Nyhan syndrome), or a protein/enzyme with an altered charge and, consequently, changed mobility when subjected to electrophoresis (*e.g.*, HbS produced by persons heterozygous for the sickle cell gene). Analysis of family pedigree also provides information of the likelihood of an individual being carrier for a genetic disease.

Once genotypes of both the prospective parents become known, it is a simple matter to work out the probability of their child inheriting the disease. If parents heterozygous for a genetic disease decide to produce a child, it is now possible, through appropriate tests done about 6-8 weeks after conception to advise them if their child has inherited the disease. This is achieved by obtaining foetal cells from biopsies of trophoblastic villi, which are an external part of human embryo and later form a part of the placenta. Foetal cells may also be obtained by amniocentesis done usually after 18 weeks of pregnancy. The cultured foetal cells may be used for determining their karyotype, levels of the critical enzymes and restriction enzyme digestion patterns of DNA. Such an ***antenatal diagnosis*** is now feasible for more

than 35 genetic diseases and for a variety of chromosomal defects. The purpose of such a diagnosis is premature termination of abnormal foetuses. This approach has helped to reduce the incidence of *thallassaemia* from 30 to 2 per year in the Cypriot community in England. However, antenatal diagnosis may sometimes be misused, *e.g.*, it may be used to selectively abort the foetuses of one sex. There were some newspaper reports that some parents in India are using amniocentesis to selectively abort female foetuses, obviously to save themselves from the hardships and sufferings they must face to arrange the exceedingly high amounts of dowry for their daughters. If such a practice becomes widespread, its consequences for the society should be obvious to even the most dimwitted.

Genetic counselling and antenatal diagnosis provide definite relief to the possible parents and reduce the frequency of genetically defective individuals in the population. However, it is unlikely that these measures would eliminate the deleterious alleles from a population. This is so because most genetic defects are recessive and heterozygotes for such alleles are normal. Thus even after a total ban on reproduction by the homozygotes for such recessive alleles they would remain in the population through the heterozygotes; therefore, even such an extreme selection would lead to only a slow decline in their frequency. Further, it is not likely that all the couples in any society will willingly submit themselves, at least in the foreseeable near future, to these procedures. But genetic counselling has become a routine aspect of medical practice in most developed countries.

43.4.3.4. Gene Therapy

Human beings suffer from more than 5000 different diseases caused by single gene mutations, *e.g.*, cystic fibrosis, acatalasia, Huntington's chorea, Tay sach's disease, Lesch-Nyhan syndrome, sickle cell anemia, mitral stenonsis, Hunter's syndrome, haemophilia, several forms of muscular dystrophy etc. In addition, many common disorders like cancer, hypertension, atherosclerosis and mental illness seem to have genetic components. Malignant cells may arise due to mutations in two types of genes, *viz.*, oncogenes and tumour suppressor genes; both the types of mutant alleles are involved in malignant transformation of cells.

Gene therapy may be defined in broad general terms as follows: Introduction of a normal functional gene into cells, which contain the defective allele of concerned gene with the objective of correcting a genetic disorder or an acquired disorder. Application of gene therapy involves the following basic developments in genetics, molecular biology and biotechnology: (*i*) identification of the gene that plays the key role in the development of a genetic disorder, (*ii*) determination of the role of its product in health and disease, (*iii*) isolation and cloning of the gene, and (*iv*) development of an approach for gene therapy.

The candidate disorders for gene therapy are selected on the basis of the following criteria: (*i*) the disease should be life-threatening, (*ii*) the gene responsible for the disease has been cloned, (*iii*) a precise regulation of the gene should not be required, and (*iv*) a suitable gene delivery system should be available.

43.4.3.5. Types of Gene Therapy

Gene therapy may be classified into two types: (*i*) germ line gene therapy and (*ii*) somatic cell gene therapy. In case of ***germ line gene therapy***, germ cells, *i.e.*, sperms or eggs (even zygotes), are modified by the introduction of functional genes,

which are ordinarily integrated into their genomes. Therefore, the change due to therapy is heritable and passed on to later generations. This approach, theoretically, is highly effective in counteracting the genetic disorders. However, this option is not considered, atleast for the present, for application in human beings for a variety of technical and ethical reasons.

In ***somatic cell gene therapy*** the gene is introduced only in somatic cells, especially of those tissues in which expression of the concerned gene is critical for health. Expression of the introduced gene relieves/eliminates symptoms of the disorder, but this effect is not heritable as it does not involve the germline. Somatic cell therapy is the only feasible option, and clinical trials have already started mostly for the treatment of cancer and blood disorders. This approach is divided into two groups on the basis of the end result of the process: (*i*) addition or augmentation gene therapy and (*ii*) targetted gene transfer.

43.4.3.6. Augmentation Therapy

In this type of somatic cell gene therapy, the functional gene is introduced in addition to the defective gene endogenous to the cell(s), *i.e.*, the modified cells contain both the defective (endogenous) as well as the normal (introduced) copies of the gene. There are two general approaches to augmentation therapy. The *first approach* was used in the first two patients on whom gene therapy was attempted to correct the genetic disorder called Severe Combined Immune Deficiency (SCID) syndrome produced by adenosine deaminase (ADA) deficiency. (*i*) Normal ADA gene copies were produced by cloning and then (*ii*) packed into a defective retrovirus; most of the viral genes were replaced by the ADA gene. (*iii*) Lymphocytes were isolated from the patients, and (*iv*) the recombinant retroviruses were used to infect the lymphocytes. Finally, (*v*) the infected cells expressing the ADA gene were injected back into the patients. The normal ADA gene was expressed in the patients, and ADA deficiency was partially corrected; this resulted in an improvement in the patient's immune system.

A variety of viral vectors have been used to deliver genes into target/stem cells, *e.g.*, lymphocytes, bone marrow cells, cultured *in vitro*. The stem cells themselves are obtained either from the concerned patient (more desirable) or from a matched donor. The reservations about safety of retroviral vectors is sought to be solved by developing suicide vectors, which can not replicate after delivery of the gene. The other main problems of this approach are: (*i*) low frequency of transfection of stem cells, (*ii*) stability of the integrated gene, (*iii*) duration of the gene expression, (*iv*) lack of proper regulation of gene expression etc. More recently, interest has focussed on physical methods of gene delivery like Ca^{2+} phosphate coprecipitation, particle gun, electroporation etc.

The *second approach* is the direct injection of DNA into the tissues either as protein complexes (in order to bring about receptor-mediated transfer of DNA into a specific tissue, *e.g.*, liver) or even as naked DNA into muscle or skin. Interestingly, these cells take up the DNA and express the gene product. Exciting results have been obtained with experimental familial hypercholesterolemia, where LDL receptor levels have been augmented by injection of the gene as a sialoglycoprotein complex. The problems, in this approach as well, relate to the frequency of cells taking up and expressing the gene and, more particularly, the duration of expression.

The gene delivery methods used for gene therapy can also be used for treatment

of cancer or AIDS. In case of cancer, a toxin encoding gene can be delivered into the cancer cells. Similarly, appropriate interleukin genes can be delivered to boost the body's defence mechanisms in the case of AIDS.

43.4.3.7. Targetted Gene Transfer

Targetted gene transfer or ***gene targetting*** uses homologous recombination to replace the endogenous gene with the functional introduced gene. The first case of such gene transfer (in 1985) was used to disrupt the human β-globin gene in cultured cells. Subsequently, over 100 mammalian genes have been modified by this approach. Gene targetting can be used either to inactivate (by disruption) a functional endogenous gene or to correct a defective one (*this is the aim in gene therapy*). The initial gene targetting to disrupt the human β-globin gene used a double selection strategy called ***positive-negative selection***.

The vectors employed for gene targetting are of two types: (*i*) insertion vectors and (*ii*) replacement vectors. The ***insertion vector*** is linearized by restriction cleavage within the sequence to be targetted; the targetted sequence provides the site for recombination and is different from the gene to be introduced. Hence the sequence to be introduced is located in the inner region of the vector and is flanked by the sequences involved in recombination. A recombination of such a vector with its homologous cellular sequences produces a duplication of the targetted sequence; this is called *insertional recombination*. In contrast, a linearized ***replacement vector*** has the two halves of the target gene at its two ends. Recombination occurs within the two halves of the target gene, replacing a portion of the endogenous gene sequence by that of the introduced gene; this is called ***replacement recombination***. There is no duplication of sequences, and the target gene becomes disrupted. Therefore, this approach can not be used for gene therapy.

A strategy has been devised to modify only a small sequence of the target gene without the attendant gene duplication/disruption produced by insertional/replacement recombination. This approach, called ***in-out method of gene targetting***, consists of the following two steps.

1. The first step, called ***'in' step***, is targetted gene transfer using an insertion vector; the appropriately targetted cell will have a gene duplication.
2. The second step, termed as ***'out' step*** , depends on either intrachromosomal recombination (between the introduced and the endogenous genes) or unequal sister chromatid exchange between homologous chromosomes. The recombination product of interest is a chromosome, which has only a single and functional copy of the introduced gene.

The in-out strategy has been tested using the HGPRT (hypoxanthine-guanine phosphoribosyl transferase) gene. The gene was targetted into a mouse embryonic stem cell line; subsequently, it has been successfully used with some other genes. This procedure is ideal for gene therapy.

Gene targetting is the strategy of choice for gene therapy for the following reasons. (*i*) The targetted gene is changed in a precise and specific manner. (*ii*) The introduced functional gene is placed in the same context, *i.e.*, it is flanked by the same DNA sequences, as the replaced endogenous gene. And (*iii*) no other gene of the genome is affected. The major limitation of the approach is the low frequency of homologous recombination; this problem, however, is being removed by

refinements of the technique. The feasibility of gene targetting has been demonstrated in a number of different cell types for several different genes. It is expected that targetted therapy would become a feasibility for many genetic diseases in the near future.

The various proposals on eugenics involve a number of religious, social, moral and political issues, which have to be discussed and resolved. The issues involved in genetic surgery and in negative eugenics are relatively more straight forward than those involved in positive eugenics. For example, the concept of 'superior' type is likely to vary from one society to another, even from individual to individual in a single society, and, more importantly, over a period of time in the same society as the social concepts and values are bound to change with time.

Biotechnology

- Use of organisms/their components to generate products/services.
- *Fermentation technology* : Microbes used to obtain products.
- *Plant cell cultures* : Mainly to obtain improved crops.
- Gene transfer in plants using Ti plasmid.
- Transgenic crops now in cultivation in several countries.
- Monoclonal antibodies.
- DNA/RNA probes.
- Recombinant/synthetic vaccines.
- Protein drugs produced in bacteria/plants ('hirudin').
- Micro-organisms used for metal extraction from ores, waste treatment, etc.

43.5. BIOTECHNOLOGY

Biotechnology consists of a variety of techniques, more specifically laboratory techniques, designed to genetically improve and/or exploit living systems or their components for the benefit of man. Gibbs and Greenhalgh (1983) have defined biotechnology as "*the use of living organisms in systems or processes for the manufacture of useful products; it may involve algae, bacteria, fungi, yeasts, cells of higher plants and animals or subsystems of any of these or isolated components from such living matter*".

In its simplest form, it employs microorganisms to change simple organic molecules (*e.g.*, sugars etc.) into more useful products (*e.g.*, alcohol, acetic acid etc.) and to produce some unique biochemicals (*e.g.*, antibiotics). This form of biotechnology is relatively older and is widely known as ***fermentation technology.*** On the other extreme of the spectrum are ranged the sophisticated techniques of recombinant DNA technology (Chapters 34-37) which hold the promise of revolutionising industrial, medical and genetic concepts and processes.

QUESTIONS

1. "Man suffers from many genetic diseases'. Discuss the validity of this statement.
2. Briefly describe the various measures suggested for eugenics, and discuss the usefulness of such measures.
3. 'Man is making every possible attempt to control the direction and the pace of evolution of living organisms'. Discuss the above statement.

4. Explain the various causes of cancer.
5. Explain the functions of protooncogenes, and describe briefly the mechanisms by which they become oncogenes.
6. Write short notes on the following:

(*i*)	genetic disease,	(*ii*)	structural chromosomal aberrations in man,
(*iii*)	euphenics,	(*iv*)	eutelegenics,
(*v*)	genetic surgery,	(*vi*)	positive eugenics,
(*vii*)	biotechnology,	(*viii*)	cancer,
(*ix*)	tumour,	(*x*)	immortalization,
(*xi*)	transformation,	(*xii*)	DNA tumour viruses,
(*xiii*)	tumour retroviruses,	(*xiv*)	tumour proteins,
(*xv*)	protooncogenes,	(*xvi*)	tumour supressor genes.

Glossary

Abundant mRNA. An mRNA present in a large number of copies per cell; in each species, a small number of different mRNAs.

Acentric Chromosome. A chromosome fragment without centromere.

Acrocentric Chromosome. A chromosome with its centromere near one end.

Active Site. In a protein (usually, enzyme) molecule; that localized site of molecule to which a substrate binds.

Additive Gene Effect. In case of polygenes; the effects of different genes affecting a character being cumulative or additive in nature.

Agglutinin. An antibody present in blood serum specific to an antigen present on red blood cells (RBCs); it causes the agglutination of RBCs having the concerned antigen. *Syn.,* antibody.

Agglutinogen. An antigen present on RBCs; it reacts with specific agglutinins to cause an agglutination of RBCs. *Syn.,* antigen.

Albinism. The absence of melanin pigment in skin, hairs, eyes etc. of an animal; the absence of chlorophyll in a green plant.

Allele. An alternative form of a gene; alleles occupy the same position or locus on the homologous chromosomes; usually, alleles do not show complementation, and yield typical monohybrid ratio in F_2.

Allelic Exclusion. In lymphocytes; expression of only one of the two alleles encoding the expressed immunoglobulin.

Allopolyploid. A polyploid having two or more distinct genomes; usually produced by chromosome doubling of interspecific hybrids (F_1s).

Allosteric Effect. The interaction of a small molecule with a protein molecule at one site modifies the activity of another site of the same protein molecule.

Allotetraploid. An allopolyploid having four genomes; usually two copies each of two genomes; may be four different genomes.

Amphidiploid. An allopolyploid having two copies each of two or more distinct genomes; natural amplidiploids behave as diploids during meiosis, *i.e.,* they show normal bivalent formation, *e.g.,* wheat, tobacco, oats, cotton etc.

Amber Codon. Codon UAG; chain termination codon.

Aminoacyl-tRNA. A tRNA carrying an amino acid; the $—NH_2$ group of the amino acid is covalently linked to either 3′— or the 2′ —OH of the terminal base of tRNA.

Aminoacyl-tRNA Synthetase. The enzyme that links an amino acid specifically to that tRNA, which has anticodon for that amino acid.

Amplification. Production of additional copies of a chromosomal sequence; amplified sequence may be pressent as chromosomal or extrachromosomal DNA.

Anaphase. The stage of cell division following metaphase; homologous chromosomes (A I of meiosis) or sister chromatids (A II of meiosis and in mitosis) begin to move from the equatorial plate towards the opposite poles.

Aneuploid. An organism or a cell having one or few chromosomes more or less than the normal disomic number of the species; the chromosome number is not a multiple of the basic number (x).

Annealing. Pairing between complementary strands of DNA to yield a double-stranded molecule.

Antibody. A protein molecule that interacts specifically with an antigen molecule; usually produced by an animal in response to the antigen in question.

Anticodon. The three bases of a tRNA that are complementary to the three bases of mRNA codon coding for an amino acid.

Antigen. A substance, usually a protein, which reacts with an antibody; usually, it induces specific antibody production when introduced into a living system, usually an animal, in which it does not occur naturally.

Antisense Strand. The strand of a DNA duplex that is transcribed; the strand complementary to mRNA.

Apomixs. The development of an embryo (and seed) without fertilization.

Asexual Reproduction. The progeny are produced through a process, which does not involve the fusion between male and female gametes.

Asynapsis. The failure, partial or complete, of pairing (synapsis) between homologous chromosomes during meiosis (zygotene of prophase I).

Attenuation. A mechanism of termination of transcription in some bacterial operons; transcription is terminated if translation of the mRNA being transcribed proceeds beyond certain sites in the mRNA.

Attenuator. The terminator sequence at which attenuation occurs.

Autogamy. Self-fertilization; union between male and female gametes produced by the same individual; in *Paramecium,* sexual reproduction within a single animal without conjugation.

Autogenous Control. The phenomenon of a gene product either inhibiting (*negative autogenous control*) or activating (*positive autogenous control*) the expression of the gene encoding the product.

Autonomous Controlling Elment. A controlling element that encodes active transposase; hence it has the ability to transpose. (cf., *nonautonomous controlling element.*)

Autopolyploid. An individual having more than two copies of a single genome.

Autoradiograph. A photograph prepared by labelling a substance with a radioactive material, and then allowing the radioactive decay in the substance to produce an image of the substance on a photographic film.

Autosome. Chromosomes other than sex chromosomes; their kind and number is the same in males and females of a species.

Auxotroph. A mutant, which is incapable of synthesizing an essential biochemical that is produced by the wild type individuals of the species.

Backcross. The cross of an F_1 hybrid to one of its parents.

Bacteriophage. A virus that infects a bacterium; usually the host cells are lysed by most bacteriophages.

Balanced Lethals. Lethal genes linked in repulsion phase; they are maintained in this phase either due to tight linkage or crossover suppression; only their heterozygotes survive.

Balanced Polymorphism. Two or more alleles of a gene maintained in the same population.

Base Analogue. Compounds with slight differences from the natural purine or pyrimidine bases; they are incorporated into DNA and RNA.

Base Pair. In DNA (or RNA) duplex; the structure formed by hydrogen bonding between A and T, or between G and C.

Basic Number. The haploid or gametic chromosome number of a diploid species (***genomic number***); in case of polyploid species, haploid number of the parental diploid species; represented by x.

Binomial Expansion. Exponential multiplication of an expression consisting of two terms joined by a + or a – sign, *e.g.,* $(a + b)^n$ or $(a - b)^n$.

Biochemical Genetics. The genetics of differences in the ability or inability to synthesize or metabolize certain biochemicals.

Biometry. The application of concepts and methods of statistics to the study of biological problems.

Biotype. Strains of a species of pathogen, particularly an insect pest, differing in their ability to attack different varieties of the same host species. *Syn.,* physiological race.

Bivalent. The structure formed by pairing of two homologous chromosomes during meiosis; seen during pachytene onward till the beginning of anaphase I.

CAAT Box. In eukaryotic transcription units; a conserved sequence located upstream of strartpoint, and recognized by a large group of transcription factors.

Cap. The structure added to the 5′-end of eukaryotic mRNAs, following transcription; the terminal phosphate of 5′ GTP is linked to the terminal base of mRNA.

Carrier. An individual heterozygous for a recessive, usually detrimental, allele of a gene.

Caryokinesis. The division of a nucleus into two daughter nuclei; through mitosis or meiosis.

Catabolite Repression. Decreased expression of many bacterial operons as a result of addition of glucose; caused by a decrease in the level of cyclic-AMP.

cDNA. A DNA strand (or molecule) produced by reverse transcription of an RNA molecule.

Cell cycle. The period from one cell division to the next.

Cell hybrid. A cell produced by fusion of two somatic cells; its nucleus has chromosome complements of both the fusion partners.

Centromere. The ordinarily localised region of a chromosome where spindle fibers attach; involved in chromosome movement during cell division; seen as a constriction in each mitotic metaphase chromosome.

Character. Derived from characteristic; a morphological, anatomical, physiological or behavioral feature of an organism; usually, the end-product of development; a result of the genotype, environment and genotype × environment interaction effects.

Chiasma. The point of contact (initially, representing the point of exchange) between nonsister chromatids of homologous chromosomes during and after diplotene (till the beginning of anaphase I); produced as a result of crossing over.

Chimera. An organism or organ having tissues of different genetic constitutions or origin; a consequence of gene mutation, natural or artificial fusion of different zygotes or organs *e.g.*, in grafting.

Chi-Square (χ^2) Test. A statistical test for determining whether observed data agree with an expected ratio; $\chi^2 = \Sigma[(O - E)^2/E]$.

Chloroplast. A highly differentiated plastid with typical granna; contains all the chlorophyll in higher plants; has DNA and participates in cytoplasmic inheritance.

Chromatid. Of a metaphase chromosome, one of the two identical longitudinal strands, produced by replication of the single chromatid of a telophase chromosome; the two chromatids of a chromosome are called ***sister chromatids.***

Chromatin. A complex between DNA, histones and other proteins, making up the chromosomes; in form of 300 Å ***chromatin fibers*** (derived from, chromaticity = property of taking up stain).

Chromocentre. In salivary gland chromosomes of certain Diptera; the body produced by a fusion of the heterochromatic regions of the autosomes and *X* chromosome, and the entire *Y* chromosome.

Chromomere. Small bead-like structures present in chromosomes, especially observable during prophase I of meiosis, having characteristic size and location; most likely produced by a more dense packing of chromatin.

Chromosomal Aberration. A change in chromosome number from the diploid state, or a change in chromosome structure from the normal chromosome complement.

Chromosome. Nucleoprotein bodies composed of 300 Å chromatin fibers; clearly observable during cell division; dark-staining with basic dyes; bearers of the genetic material and the genes.

Chromosome Banding. The staining of chromosomes following specific procedures, which produce many dark and light zones along their lengths; the pattern of banding is highly specific, and all the 23 human chromosomes can be identified.

Chromosome Walking. A sequential isolation of clones carrying overlapping sequences of DNA, allowing large regions of chromosome to be covered, usually, to reach a locus of interest.

cis-*acting Locus*. A DNA sequence that affects the activity of only that DNA sequence to which it is covalently linked; sometimes called element.

cis-*Arrangement.* See, ***Coupling Phase.***

cis/trans *Test.* Analysis of the effects of two mutations present in *cis* and *trans* configuration on the expression of the concerned gene.

Cistron. A segment of DNA, which codes for one polypeptide; the unit of function identified by *cis-trans* or complementation test; usually, mutants within a cistron do not show complementation; generally, *accepted as a synonym for gene.*

Clone. All the individuals obtained from a single individual through asexual reproduction *i.e.*, produced by mitotic cell division.

Coding strand. The strand of DNA duplex that has the same base sequence as mRNA. *Syn.*, sense strand.

Codominance. Both the alleles of a gene express themselves in the heterozygotes.

Codon. A group of three adjacent nucleotides of mRNA, which code for one amino acid; ***nonsense codons*** specify no amino acid and function in polypeptide chain termination.

Coefficient of Coincidence. In case of linkage; the ratio of observed frequency of double crossovers to their expected frequency.

Coefficient of Interference. One minus the value of coefficient of coincidence; in case of linkage.

Colchicine. An alkaloid derived from autumn crocus; it prevents the organisation of spindle apparatus and, thereby, induces chromosome doubling.

Colinearity. The sequences of mutant alleles in the linkage map of a gene being identical with the sequences of altered amino acids generated by these alleles in the polypeptide specified by the gene in question.

Colony Hybridization. *In situ* hybridization to identify bactria carrying recombinant vectors that have DNA insert homologous to the probe used for hybridization.

Colour Blindness. Inability of an individual to distinguish between two colours, *e.g.*, red-green colour blindness; usually, sex-linked.

Competence. The ability of a bacterial cell to take-up exogenous DNA and, as a result, to become transformed.

Complementary Genes. Two (or more) dominant genes capable of producing a character only when both (or all) of them are present together in the dominant state.

Complementation. The F_1 obtained by crossing two mutant strains has the wild type phenotype.

Complementation Test. A test to determine, on the basis of complementation, if two mutant alleles are located in the same gene; if there is complementation, the mutants are located in different genes; otherwise they are located in the same gene; cases of intragenic complementation known, especially in eukaryotes.

Complex Locus. In *Drosophila*, a very large locus whose properties are consistent with the function of a single protein.

Composite Transposon. A transposon having a central region flanked on each side by IS elements; transposase is encoded by either or both of the IS elements.

Concerted Evolution. Ability of two genes to evolve together as if they were part of the same locus.

Concordance. The two members of a pair, *e.g.*, identical twins, showing a similar phenotypic expression for a trait.

Conditional Lethal. A lethal gene that kills the cells under certain conditions (nonpermissive conditions) but not others (permissive conditions).

Conjugation. The union between individuals in unicellular organisms; individuals of different mating types either fuse completely (*Chlamydomonas*), exchange nuclei (*Paramecium*), or may show one-way-transfer of the genetic materi l (bacteria, *e.g.*, *E. coli*).

Consensus Sequence. An idealized sequence in which each position is occupied by the base most often found in the actual sequences analyzed.

Constitutive Enzyme. An enzyme, which is produced at a constant rate whether it is needed by the organism or not.

Constitutive Heterochromatin. The inert state of permanently non-expressed sequences; usually, satellite DNA.

Continuous Variation. Variation not separable into distinct classes; governed by polygenes; markedly affected by the environment; in quantitative traits; analysis based on measurement data.

Controlling Elements. First discovered in maize; transposable elements, *e.g.*, *Ac-Ds*; autonomous (*Ac*) or nonautonomous (*Ds*).

Corepressor. A small molecule that binds a repressor protein making the latter active and capable of binding to operator, thereby, repressing transcription; repressor is inactive without the corepressor.

Cosmid. A plasmid having the *cos* site of λ phage.

Coupling Phase. In linkage; the dominant alleles of two or more genes present in the same chromosome and, hence, linked together; dominant alleles are contributed by the same parent.

Crossing Over. An exchange of homologous segments between the nonsister chromatids of homologous chromosomes leading to recombination between linked genes; inferred on the basis of recombination between linked genes; demonstrated cytologically; occurs during pachytene.

Crossover Unit. The distance between two linked genes, which permits 1% recombination between them. *Syn.*, map unit, centiMorgan.

Cyclin. The regulatory subunit of M phase kinase; accumulates throughout the cell cycle, and is destroyed by proteolysis after onset of mitosis.

Cytogenetics. A branch of biology devoted to the study of chromosomes and their implications in genetics.

Cytokinesis. The division of cytoplasm into two or more parts; usually follows caryokinesis or nuclear division.

Cytology. A study of the structural and functional organisation of the cell.

Cytoplasm. The entire content of a cell other than the nucleus; includes ER, plastids, mitochondria, microtubules etc., and the hyaloplasm.

Cytoplasmic Inheritance. Inheritance due to the genes located in the cytoplasm (***plasma genes***), *e.g.*, mitochondria and plastids; typical uniparental transmission and lack of segregation; in higher plants and animals, transmission through the female parent *Syn.*, extranuclear, extrachromosomal and maternal inheritance.

Deficiency. The loss of a segment from a chromosome; deficiency heterozygotes are hemizygous for the genes located in the deleted segment; many deficiencies produce genetic effects similar to gene mutations. *Syn.*, deletion.

Degenerate Code. One amino acid is specified by more than one codon.

Degrees of Freedom. In χ^2-test; the number of classes to which an arbitrary number of observations can be assigned to yield the given total; ordinarily it is one less than the number of classes in the data.

Denaturation. The loss of natural or native configuration of a macromolecule; generally, accompanied with a loss in the biological activity of the molecule; a

separation of the two strands of DNA double-helix producing single-stranded DNA molecules.

de novo. Arising afresh.

Determination. The process due to which an embryonic cell becomes committed to differentiate only into a specific adult organ.

Diakinesis. A substage of prophase I of meiosis; the chiasma terminalization is complete; the bivalents are short and thick, and well-spread in the cell.

Dicentric Chromosome. A chromosome having two centromeres.

Differentiation. The modification of identical cells derived from a zygote into the different cell types; usually associated with organ formation, and characterised by differences in the kinds of genes being transcribed.

Dihybrid. The progeny from a cross between two homozygous parents differing for two genes (or characters); an individual heterozygous for two genes.

Dioecious. Plant species in which unisexual (male and female) flowers occur on different plants; in such species, plants are either male or female.

Diploid. A cell, organism or species having two copies of a single genome; specified by $2x$ (cf. $2n$).

Diplotene. The stage after pachytene and before diakinesis during prophase I of meiosis; characterised by the movement of homologous chromosomes, paired to form bivalents, away from each other and the appearance of chiasmata.

Discontinuous Variation. Variation, which can be divided into distinct classes; in case of qualitative characters; usually little affected by the environment; yields typical F_2 ratios.

Discordant. When two members of a pair, *e.g.*, twins, show dissimilar phenotypes.

Disjunction. The separation of homologous chromosomes (AI of meiosis) or sister chromatids (A II of meiosis and anaphase of mitosis), and their movement to the opposite poles.

Dizygotic Twins. Twins originating from two separate eggs, each fertilized by a separate sperm; may be of the same or different sex. *Syn.*, fraternal twins.

DNA. Deoxyribonucleic acid; the genetic material; serves as the template for DNA replication and transcription; may be unique or repetitive.

DNAase. An enzyme that specifically digests or hydrolyses DNA.

DNA Insert. The forein DNA segment integrated into a vector.

DNA Polymerase. The enzyme that synthesizes DNA using DNA as template; responsible for DNA replication and repair.

DNA Replicase. DNA polymerase responsible for DNA replication.

Dominance. In the heterozygote, one allele expresses itself at the expense of the other.

Dosage Compensation. The phenomenon and the mechanisms by which a single X chromosome in one sex, *e.g.*, males in *Drosophila* and mammals, functions at the same level as do two X chromosomes in the other sex.

Double Cross. A cross between two single crosses, *i.e.*, between two distinct F_1s, each F_1 being derived from two different parents.

Double Crossover. In linkage; the recombinant type produced by simultaneous crossing over on both the sides of a gene; it is the least frequent class in the test cross progeny of a three-point cross.

Drift. See, ***Random genetic drift.***

Duplicate Genes. Two dominant genes produce the same phenotypic effect whether they are alone or together, the alternative phenotype being produced only when both are in the recessive state.

Duplication. The occurrence of a chromosome segment more than once in the same chromosome or in a genome.

Ectopic Expression. Expression of a gene in a tissue in which it is not usually expressed.

Egg (Ovum). The gamete produced by a female organism.

Element. A sequence of DNA that serves as the binding site for a regulatory protein, *e.g.*, operator, promoter etc.; they are *cis*-acting.

Embryo. The early stages of development of an organism from the zygote.

Embryo Sac. In plants; the structure derived from a megaspore through mitosis; generally, contains one egg cell, two synergids, three antipodal cells and two polar nuclei.

Endogenote. In bacteria; the chromosome segment, which is homologous to the exogenote (the DNA segment transferred from a donor cell to the cell in question).

Endomitosis. Two or more chromosome replications not followed by nuclear or cell division; leads to endopolyploidy, *i.e.*, each replication doubles the existing chromosome number of the cell.

Endonuclease. An enzyme, which induces internal cuts or cleavages in a DNA molecule.

Endoplasmic Reticulum. A network of membranes in the cytoplasm; consists of tubules, vesicles and cisterns; provides the attachment site for ribosomes; may be rough or smooth.

Endopolyploidy. The condition of cells with twice, four-times etc. chromosomes of the normal somatic cells of an organism; due to endomitosis.

Endosperm. The nutritive tissue of developing seeds derived from repeated mitotic divisions of the fusion product of the two polar nuclei of an embryo sac with a sperm; usually it is $3n$, but in some cases it is $5n$.

Enhancer. A *cis*-acting element that increases the efficiency or activity of any promoter located in its vicinity; functions in either orientation; located upto several kb upstream or downstream of the affected promoter.

Environment. The sum total of all living and nonliving factors outside an organism, which affect the development and the functions of the organism.

Enzyme. A protein, which catalyses or speeds up a specific chemical reaction; usually in a living system.

Epigenetic Change. Changes that affect phenotype but not the genetype; these changes may be heritable; in some cases, due to a protein assuming an alternative conformation, *e.g.*, prions.

Episome. In bacteria; a DNA molecule, other than the bacterial chromosome, capable of replication; may be integrated into or independent of the bacterial chromosome; it is not essential for the bacterial cells, except under specific environments.

Epistasis. The expression of a gene is modified by another gene; in quantitative genetics, nonallelic interaction or simply interaction.

Equational Division. the second division of meiosis when the sister chromatids separate; when crossing over occurs between a gene and its centromere, segregation for the gene takes place during this division.

Equatorial Plate. It is the arbitrary plane at which the chromosomes are arranged during metaphase.

Erythroblastosis Foetalis. The haemolytic anemia of newborn infants; when mother is Rh^- and father is Rh^+, Rh^+ infants may develop this condition.

Estrogen. The female hormone.

Euchromatin. The parts of a chromosome, which take deep stain during cell division but do not stain during interphase; genetically active parts of chromosomes.

Eugenics. The application of principles of genetics for the improvement of mankind.

Eukaryote. An organism whose cells have true nuclei, *i.e.*, nuclei enclosed by typical nuclear envelope.

Euploid. A cell or an organism in which the chromosome number is an exact multiple of the basic number (x) for the concerned species.

Evolution. The development of new varieties, races, species, genera etc. through the origin of variation and natural selection.

Excision-repair. The sequence of one strand of DNA that contains the damage is removed and the gap so created is filled by DNA synthesis using the intact strand as template.

Exogenote. In bacteria; the chromosomal segment transferred from a donor cell to a recipient cell and homologous to a chromosomal segment of the latter (*cf.* endogenote).

Exon. The expressed sequence of an eukaryotic gene; the part of an eukaryotic gene, which is represented in the mature mRNA and in the polypeptide specified by the gene.

Exonuclease. An enzyme, which digests DNA or RNA molecules beginning at their ends.

Expected Ratio. In chi-square test; the ratio in which the data are expected to occur in a given experiment; based on known facts and theoretical considerations.

Expression Vector. A cloning vector that is designed in such a way that a coding sequence integrated into it at a specific site will be transcribed and translated into a protein.

Expressivity. The degree of phenotypic expression of a gene in the different individuals; it may be uniform or variable.

F_1. The progeny obtained by crossing or mating two strains or individuals; first filial generation.

F_2. The second filial generation; obtained by self-fertilization of or intermating among the individuals of F_1 generation.

F Factor. In bacteria; fertility factor, which converts a cell into male; cells with F factors have characteristic F-pili.

Faculative Heterochromatin. Heterochromatization of such sequences that are otherwise euchromatic; usually, due to their location in the vicinity of constitutive heterochromatin.

Feedback Inhibition. Inhibition of an enzyme of a biosynthetic pathway by the end-product of the pathway when it accumulates in the cell upto the threshold concentration; usually the first enzyme of the pathway is involved.

Fertility. The ability of an organism to produce progeny.

Fitness. The number of progeny produced by an individual or a genotype in relation to the average of the population or to that of the best fit genotype.

Fixation. The frequency of an allele in a population attaining the value of one.

Footprinting of DNA. The technique of identifying the site in a DNA duplex to which a protein is bound; binding of the protein protects the site from nuclease action.

Frame-shift Mutation Mutation produced by deletion or addition of bases that are not in multiples of three; as a result, reading frame changes downstream of the site of addition or deletion.

Founder Principle. The theory that small isolated populations diverge, in terms of gene and genotype frequencies, from the parent population due to sampling error at the time of the establishment of these populations.

G1 Phase. In eukaryotic cell cycle; the period after telophase of the previous division and the start of S phase, *i.e.*, DNA replication.

G2 Phase. In eukaryotic cell cycle; the period after S phase and the start of following prophase.

Gamete. A specialized haploid (usually) cell, which participates in sexual reproduction; one male and one female gametes fuse to produce a zygote.

Gametic Chromosome Number. The chromosome number normally present in the gametes of a species; ordinarily, it is the haploid number (n).

Gametogenesis. The production of male and female gametes.

Gene. A segment of DNA, which specifies a single polypeptide or an RNA molecule *e.g.*, rRNA and tRNA (***structural gene***); they are trans-acting.

Gene Conversion. A process of nonreciprocal recombination due to which an allele is present in an excess of the other in the recombination products.

Gene Family. A set of genes whose exons are related; derived from an ancestral gene by duplication and variation due to mutation.

Gene Frequency. The proportion of one allele relative to all the alleles of a gene in a Mendelian population.

Gene Pool. The sum total of all the alleles present in the breeding or reproductive members of a random mating population.

Gene Symbol. A letter or a group of letters used to denote a gene; usually derived from the name of the mutant phenotype; the first letter of the symbol for the dominant allele is capital, while all the letters in the symbols for recessive alleles are in lower case.

Genetic Block. The block in a biochemical reaction due to a mutant allele of the concerned gene; the concerned enzyme produced by the mutant allele is defective or nonfunctional.

Genetic Code. The correspondance between triplet codons in mRNA and amino acids in protein.

Genetic Drift. See, ***Random Genetic Drift.***

Genetic Load. The proportion by which the fitness of a random mating population decreases as compared to that of the optimum genotype due to the presence of deleterious alleles in the population.

Genetic Marker. An allele that produces an easily detectable phenotype that can be easily followed in segregating generations.

Genetics. The branch of biology concerned with the study of heredity and variation.

Genome. A complete set of chromosomes of a diploid species; all the members of a genome are distinct from each other in gene content and, often, in morphology; members of a genome do not pair.

Genotype. The genetic constitution of an organism.

Genetic Fine Structure. Essentially a linkage map of the various mutant alleles of a gene; based on complementation and recombination data.

Germ Cell. A cell destined to produce gametes for sexual reproduction.

Germplasm. The hereditary material transmitted from parents to offspring through germ cells.

Gonad. The organ concerned with the production of gametes, *e.g.*, ovary and testis; in case of animals.

GT-AG Rule. As a rule, the first two nucleotides of nuclear introns are GT, while the last two nucleotides are AG.

Gynandromorph. An individual in which a part of the body is male, while the rest of it is female; usually in *Drosophila;* mosaics for the sex chromosome; male part (*XO*) may range from a small section of the body to half of the body.

Gyrase. In *E. coli.*, a type II topoisomerase that introduces negative supercoils into DNA.

Haploid. A cell or an organism having the gametic chromosome number.

Haplotype. The combination of alleles in a given region of a chromosome; in effect, the genotype for a specific region of a chromosome.

Hapten. A small molecule that acts as an immunogen, when linked to a protein.

Hardy-Weinberg Law. In random mating populations, gene and genotype frequencies remain constant generation after generation provided there is no migration, mutation, selection or random drift.

Hardy-Weinberg Equilibrium. In a random mating population, a gene with two alleles exhibits $p^2\ AA : 2pq\ Aa : q^2\ aa$ genotypic frequencies (where, $p + q = 1$).

Helper Virus. It provides the functions missing from a defective virus, *e.g.*, a tumour causing retrovirus, so that the latter is able to complete the infective cycle when there is a mixed infection of a host cell.

Hemizygous. A diploid organism having only one copy of a gene in place of the normal two, *e.g.*, in case of sex linkage and in deletion heterozygotes.

Hemophila. The tendency of bleeding freely from even minor cuts due to an inability of the blood to clot; a genetic disease due to a sex-linked recessive gene.

Heredity. The transmission of traits (more precisely, genes) from parents to offspring.

Heritability. The ratio of genotypic variance (*broad sense heritability*) or additive genetic variance (*narrow sense heritability*) to the total phenotypic variance.

Hermaphrodite. An individual with both male and female reproductive (functional) organs.

Heterochromatin. The parts of chromosomes (or chromatin), which take up dark stain during interphase; genetically inactive; usually, contains repetitive DNA sequences, which are rarely transcribed; may be constitutive or facultative.

Heteroduplex. A DNA molecule in which the two strands are mismatched for one or more nucleotides, *e.g.*, one strand may carry a deletion for some bases.

Heterogametic Sex. The sex, which produces two types of gametes with regard to the sex chromosomes, *e.g.*, mammalian males (*XY*).

Heterogeneous Nuclear RNA (hnRNA). Consists of transcripts of nuclear genes produced by RNA polymerase II; shows a wide variation in size and low stability.

Heterokaryon. A mycelium or cell having two nuclei of different genotypes.

Heteroploid. An individual having a chromosome number other than the normal diploid (2*x*) number.

Heteropycnosis. Certain chromosomes or their parts staining differently from the rest of the chromosomes during cell cycle, *e.g.*, in the case of heterochromatin. *Adj.*, heteropycnotic.

Heterosis. The superiority of a heterozygote over both the concerned homozygotes in respect of one or more characters; in practice, the superiority of an F_1 over both its parents.

Heterozygote. An individual having two different alleles for one or more genes. *Adj.*, heterozygous; *noun*, heterozygosity.

Hexaploid. An individual having six genomes of one (*autohexaploid*) or more (*allohexaploid*) kind.

Hfr. In *E. coli; high frequency recombination;* cells in which the *F* factor is integrated into the bacterial chromosome.

Histone. The basic proteins associated with chromatin in eukaryotes.

Holandric Gene. A gene carried in the *Y* chromosome and, thereby, showing only male-to-male transmission; confined to the male sex only (in mammals).

Homeotic Genes. Those genes mutations in which convert one body part into another, *e.g.*, an insect leg may change into an antenna.

Homogametic Sex. The sex, which produces only one type of gametes with regard to the sex chromosome.

Homologous Chromosomes. The chromosomes, which pair regularly and freely with each other during meiosis; they are identical in gene content and, usually, in morphology.

Homozygote. An individual having two copies of the same allele for one or more genes under consideration. *Adj.*, homozygous; *noun*, homozygosity.

Hormone. A compound produced by the cells of one part of the body, which acts in the other parts of the organism.

Housekeeping Genes. These genes are expressed in all cells; presumed to encode proteins needed for sustenance of all cell types. *Syn.*, constitutive genes.

Hybrid. The progeny obtained by crossing or mating two individuals or strains having different genotypes.

Hybrid Dysgenesis. In *Drosophila*; the phenomenon of hybrids between certain strains being sterile; due to P elements.

Hybrid Vigour. See, ***Heterosis.***

Hybridization. The process of obtaining progeny through the union between male gametes of one strain with the female gametes of another strain. *Syn.*, crossing, mating.

Hybridoma. A cell line produced by fusing a mycloma cell (that does not produce an antibody) with a B lymphocyte; continues to divide and express the antibody produced by the B lymphocyte fusion parent.

Identical Twins. Twins derived from the misdivision of a single zygote; members of such twins have the same sex and are identical in genotype. *Syn.*, monozygotic twins.

Imaginal Disc. In insects; a group of embryonic cells determined to develop into specific adult organs; present as thickenings of epidermis in the larvae.

Immortalization. In animal cells grown *in vitro*; the phenomenon of cells becoming capable of infinite growth *in vitro*; produces permanent cell lines.

Immunogen. A molecule that induces antibody production against itself.

Immunoglobulin. A class of proteins acting as antibodies to specific antigens; insoluble in water but soluble in salt solution; present in blood serum.

Imprinting, Genetic. The change that occurs during passage through the sperm or egg so that either paternal or maternal allele is expressed in young ambryo; caused by methylation of DNA.

Inbreeding. Mating between two individuals related by descent, *i.e.*, having a common ancestor.

Incomplete Dominance. The phenotypic expression of heterozygotes for a gene being intermediate between those of the two concerned homozygotes.

Independent Assortment. The segregation of one pair of alleles being independent of that of another pair of alleles, that is, an allele of one gene passes with either one of the two alleles of the other gene with equal frequency.

Inducer. A small molecule, usually a substrate, which induces the production of an enzyme, usually, concerned with the utilization of that substrate; the inducer combines with the concerned repressor to make the latter inactive and, thereby, relieves the repression of the gene producing this enzyme.

Inducible Enzyme. An enzyme, which is synthesized only when a specific inducer is present; generally, the inducer is the substrate for that enzyme.

Induction of Prophage. Excision of a prophage from host genome; occurs due to the destruction of the lysogenic repressor; excised phage causes lysis.

Inhibitory Gene. A dominant gene, which prevents the expression of another dominant gene; generally, denoted by *I*.

Insertion Sequence. See, ***IS Elements.***

In situ *Hybridization.* Cells are sequashed on a slide, their DNA is denatured, radiolabelled single-stranded DNA (or RNA) is added and allowed to hybridize with cellular DNA; the sites of hybridization are detected by autoradiography.

Interference. The phenomenon of occurrence of a crossing over at a point reducing the chances of another crossing over taking place in the neighbouring regions

of the chromosome; in the case of ***negative interference,*** the chances of a second crossing over are enhanced.

Interphase. The stage in cell cycle when the nucleus of a cell is not undergoing division; a period of active protein (G_1 and G_2 phases) and DNA (S phase) syntheses.

Intersex. A sterile organism whose secondary sex characters are intermediate between males and females of the species.

Intron. The parts of an eukaryotic gene, which are not represented in the mature mRNA molecules; derived from 'intervening sequences'.

Inversion. Rotation of a segment of a chromosome by 180° so the genes, in this segment are present in the reverse order; characteristic inversion loops are produced during meiosis in the inversion heterozygotes.

In Vitro. When biological processes are made to occur outside the organism in a vessel, *e.g.,* test tube.

In Vivo. Within the living organism.

Ionizing Radiation. A high energy radiation, which produces ions in the medium through which it passes.

IS Elements. DNA segments, of about 800-1400 bp, capable of integrating into, and dissociating from and moving to a new location in the chromosomes of prokaryotes and eukaryotes; they also induce the transposition of other *IS* elements in the same cell.

Isochromosome. A chromosome in which the two arms are identical with each other in gene content and morphology.

Isogenic Lines. Lines having identical genotypes, except for one gene at which they differ.

Isoaccepting tRNAs. Those tRNAs that have different anticodons but are charged with the same amino acid.

Isotype. A group of closely related immunoglobulins.

Kappa Particle. In *Paramecium;* a bacterial endosymbiont, which harbours a virus; it produces a toxin that kills sensitive paramecia, which do not carry the kappa particle; responsible for the 'killer trait'.

Karyotype. The chromosomes of a cell arranged according to their length; also the chromosome constitution of a cell or an individual.

Kb. Kilobase pairs (in DNA) or kilobases (in RNA); 1,000 bases or base pairs.

Kinase. An enzyme that phosphorylates (*adds a phosphate group to*) a substrate; *e.g.,* a ***protein kinase*** phosphorylates amino acids of another protein (divided into two groups in eukaryotes; tyrosine and threonine/serine kinases).

Kinetochore. See ***Centromere.***

Klinefelter's Syndrome. Sterile human males with the *XXY* chromosome constitution; other associated symptoms as well.

Lagging Strand. That strand of DNA replication of which lags behind the other; its replication is discontinuous.

Lamella. A structure formed by two membranes lying parallel to each other; usually in form of plates, vesicles or tubles.

Lampbrush Chromosome. In diplotene stage of vertebrate oocytes, particularly, in amphibia; the two chromosomes of each bivalent have numerous paired loop-like lateral extensions; loops are the sites of active transcription.

5′ Leader. The nontranslated sequence at the 5′-end of mRNA (upstream of the initiation codon).

Leader Sequence. In case of a protein; a short N-terminal sequence that is responsible for passage into or through a membrane.

Leading Strand. The strand of DNA that replicates continuously and ahead of the other strand.

Leptotene. The first substage of prophase I of meiosis; chromosomes clumped as a loose ball of wool, and are visible as fine threads.

Lethal Gene. An allele of a gene, which causes the death of all those individuals that carry it in the appropriate genotype; death occurs any time before the individuals reach adulthood.

Ligase. An enzyme, which seals the nicks or the cuts in a DNA strand.

Ligation. Formation of a phosphodiester bond to link two adjacent bases separated by a nick in one strand of a DNA duplex; also joining of both the strands of a DNA molecule with those of another DNA molecule, and joining together of two RNA molecules.

Linkage. The tendency of genes to stay together during inheritance; due to the genes being located relatively close to each other in the same chromosome; produces typical distortion of testcross ratios (of F_2 ratios as well).

Linkage Map. A linear map of the genes showing linkage with each other; it depicts the sequences in which genes are located in the chromosome as well as the frequency of recombination between the adjacent genes.

Locus. The fixed position in a chromosome occupied by a gene; often used as a synonym of gene since, ordinarily, genes occupy a fixed position in the concerned chromosome.

LTR. Long terminal repeat; a sequence directly repeated at both the ends of a retroviral genome.

Luxury Genes. Genes coding for specialized functions; expressed in particular cell types only.

Lysis. The rupture of a bacterial or host cell due to the disintegration of its cellular structures following its infection by a virus. (That is why bacterial viruses are called bacteriophage.)

Lysogenic Repressor. A protein that prevents a prophage from reentering a lytic cycle.

Lysogeny. The chromosome of a bacteriophage becoming integrated into the chromosome of its host; in case of temperate phages; such bacteria are called ***lysogenic.***

Main Band DNA. The genomic DNA that forms a broad peak on a density gradient. (*c.f.* Satellite DNA).

Male Sterility. The male gametes produced by an individual being nonfunctional.

Map Unit. See, ***Crossover Unit.***

Maternal Effect. The phenotypes of progeny being affected by the genotypes of their mothers.

Maternal Inheritance. The transmission to progeny of a character from the female parent only; see, ***Cytoplasmic Inheritance.***

Mb. Megabase; 10^6 bp of DNA or 10^6 bases of RNA.

Mean. The sum of all the observations in a sample divided by the number of observations in the sample. *Syn.*, average.

Megasporogenes. The production of female spore through meiosis in the ovules of plants.

Meiosis. Two cell divisions, occurring one after the other, following only one round of DNA replication; shows chromosome pairing and chromosome separation in the first division, and chromatid separation in the second; leads to the formation of haploid daughter cells (spores in plants; gametes in animals).

Meiotic Drive. A mechanism that leads to the production of unequal frequencies of functional gametes by a heterozygote.

Melting of DNA. See, ***Denaturation.***

Melting Temperature (Tm). The midpoint of temperature range over which a DNA molecule is denatured.

Mendelian Population. A random mating population; a group of individuals sharing the same gene pool.

Merozygote. In bacteria; a partial zygote; a cell having a segment of DNA (called ***exogenote***) homologous to a segment of its own chromosome (termed as ***endogenote***).

Messenger RNA (mRNA). The RNA molecules that carry the information for amino acid sequences of specific proteins, which is decipered with the help of ribosomes and specific tRNAs.

Metabolism. The sum total of all biochemical processes concerned with energy production and energy utilization occurring in living cells.

Metacentric Chromosome. A chromosome that has centromere in its centre or close to the centre; the two arms of such a chromosome are nearly equal.

Metafemale. A female individual, usually sterile and weak, showing a more intense expression of the secondary sex characters; due to a greater number of *X* chromosomes (*e.g., XXX*) than in the normal females (*XX*). *Syn.*, superfemale.

Metaphase. The stage during cell division after prophase and before anaphase; the chromosomes are short, thick distinct from each other and arranged on the equatorial plate.

Metastasis. Ability of tumour cells to leave the site of their origin and to migrate to other locations in the body, where they grow and form a new colony.

Microsomes. Fragmented pieces of ER associated with ribosomes.

Microsporogenesis. The production of haploid male spores (pollen grains) through meiosis in the anthers of plants.

Microtubules. The hollow, tube-like filaments formed by precisely arranged protein molecules; present as constituents of many cellular structures, *e.g.,* centrioles, flagella, spindle fibers etc.

Mitochondria. Cytoplasmic organelles enclosed by two concentric membranes, the inner membrane having characteristic infolds called cristae; concerned with oxidative phosphorylation and production of ATP; contains DNA and participates in cytoplasmic inheritance.

Mitosis. The cell division in which each nuclear division is preceded by one round of DNA replication; during anaphase sister chromatids separate and pass to the opposite poles.

Modifying Gene. A gene that increases the phenotypic expression of a major gene; usually, several modifying genes act in an additive manner and generate a continuous variation in an otherwise qualitative trait.

Molecular Chaperone. A protein that is needed for assembly or porper folding of some other protein, but which does not itself become a part of the molecule so produced.

Monocistromic mRNA. An mRNA that codes for one polypeptide; eukaryotic mRNAs.

Monoecious. In plants; a plant having both male (***staminate***) and female (***pistillate***) flowers.

Monogenic. Governed by or involving a single gene.

Monohybrid. The progeny derived by mating or crossing two individuals or strains, which differ for one gene (in practice, one character).

Monohybrid Cross. Hybridization between two individuals or strains, which differ for a single character, or which are taken to differ for a single character.

Monolayer. The one cell deep growth of animal cells *in vitro* on a solid substrate.

Monoploid (x). An individual having one copy of a single genome; haploid from a diploid ($2x$) species.

Monosomic. An individual having one chromosome less than the normal somatic complement of the species; denoted by $2n - 1$.

Monozygotic Twins. See, ***Indentical Twins.***

Morphogen. A factor that induces the development of a particular cell type in a manner that depends on it concentration.

Mosaic. An individual or tissue having cells of two or more different genotypes.

Multiple Alleles. More than two alleles of a single gene (more precisely, of a single *cistron*).

Mutable Gene. A gene having an unusually high spontaneous mutation rate.

Mutagen. A chemical or physical agent, which induces mutation.

Mutant. An organism or cell showing a mutant phenotype due to the mutant allele of a gene.

Mutation. A sudden and heritable change in the character of an organism, which does not arise due to segregation or recombination.

Mutation, Gene. A change in the base sequence of a gene; usually, produces a mutant phenotype.

Mutator Gene. A gene, which causes another gene or genes to undergo spontaneous mutation.

Muton. The unit of mutation; the smallest unit of a gene capable of undergoing mutation; represented by one nucleotide.

Myeloma. A tumour cell line derived from a lymphocyte; usually produces a single type of antibody.

Natural Selection. Differential rate of reproduction of different genotypes of an organism in response to the environmental factors; the basis of organic evolution.

Nick. In duplex DNA; absence of a phosphodiester bond between two adjacent nucleotides of one strand.

Nonautonomous Controlling Element. In eukaryotes; defective transposons that can transpose only when transposase function is provided by an autonomous element of the same type.

Nondisjunction. The failure of a chromosome pair (during A I of meiosis) or of sister chromatids of a chromosome (during A II of meiosis and anaphase of mitosis) to separate during anaphase; as a result, they pass to the same pole leading to the production of $n + 1$ and $n - 1$ gametes (meiosis) or $2n + 1$ and $2n - 1$ daughter cells (mitosis).

Nonsense codon. Codon that does not specify any amino acid; three codons: UAG, UAA, UGA; causes chain termination.

Northern blotting. Technique for transferring RNA from an agarose gel onto a nitrocellulose filter, where it can be hybridized with a complementary DNA.

Nucleic Acid. The macromolecules composed of phosphate, pentose and organic bases, *viz.*, DNA and RNA.

Nucleoid. The compact body produced by folding and supercoiling of bacterial chromosome.

Nucleolus. A relatively dense, spherical structure rich in RNA and protein, and associated with the NOR; the site of ribosome assembly and synthesis of ribosomal components.

Nucleolus Organizer Region (NOR). The segment of a chromosome containing the genes concerned with the synthesis of ribosomes; usually, seen as a constriction in metaphase chromosomes and associated with the nucleolus during prophase I of meiosis.

Nucleoprotein. A complex containing nucleic acid and protein, *e.g.*, chromatin.

Nucleosome. The bead like structure of chromatin fibers; each composed of a histone octamer and 146 base pairs of DNA.

Nucleotide. The unit of DNA and RNA organisation; each nucleotide has one molecule each of phosphate, pentose (ribose in RNA, deoxyribose in DNA) and one of the five organic bases [A, G, C, T (in DNA) and U (in the place of T in RNA)].

Nucleus. In eukaryotes; the organelle containing chromosomes and nucleolus, and surrounded by a typical nuclear envelope composed of two concentric membranes.

Nullisomic. An individual having one chromosome pair less than the normal somatic complement of the species.

Ochre Codon. The nonsense codon UAA.

Octaploid. A cell or an organism having eight identical (*autooctaploid*) or different (*allooctaploid*) genomes.

Okazaki Frament. Fragments of 1,000 to 2,000 bases produced during discontinuous replication of the lagging strand.

Oligogene. A gene producing a marked effect on the expression of a character so that the individuals can be easily grouped into separate classes on the basis of this trait; its expression is relatively little affected by the environment. (*cf.* polygenes).

Oncogene. A gene, which transforms an animal cell into a tumour forming or cancerous cell; ***cellular oncogenes*** are an integral part of the animal genome, while ***viral oncogenes*** are present in the genomes of tumour producing viruses; both are very similar.

Oocyte. The diploid cell in the ovary of a female animal, which undergoes meiosis to yield the ovum or egg cell.

Oogenesis. The formation of eggs or ova in female animals through meiosis in oocytes.

Open Reading Frame (ORF). A DNA sequence that has a continuous series of triplets coding for amino acids; potentially translatable into a protein.

Operator. In operons; the sequence to which the repressor molecule binds; it may overlap the promoter or lie adjacent to it.

Operon. A group of structural genes whose transcription is regulated by a set of a regulator gene, a promoter and an operator, which are often located upstream or onto the left of the structural genes.

Organelle. A specialized substructure of a cell with a specific structure and function; surrounded by a membrane.

Origin (of DNA replication). A sequence of DNA at which replication is initiated.

Ovary. The gonad of a female animal; in plants, the basal part of pistil containing ovules.

Ovule. In plants; the structures present in ovaries, which produce megaspores and embryo sacs and later develop into seeds.

P. The symbol denotes the parents of an individual or the parental generation.

Pachytene. A substage of prophase I of meiosis following zygotene and preceding diplotene; often n chromosomes are visible as long closely paired structures (bivalents).

Palindrome. A DNA segment in which the base-pair sequence reads the same in both the directions from a point of symmetry within the segment.

Panmictic Population. A random mating population.

Panmixis. See, ***Random Mating.***

Paracentric Inversion. An inversion in which the centromere is not included in the inverted segment.

Parthenogenesis. The development of an individual from an unfertilized egg.

PCR. Polymerase chain reaction; a technique for producing millions of copies of a DNA sequence; uses a thermostable DNA polymerase, primers complementary to the 3′-ends of the DNA segment to be amplified and the four deoxyribonucleotides; 30 or more cycles of denaturation, annealing and extension (DNA synthesis) steps.

Penetrance. The ability of a gene to express itself in all the individuals, which carry it in the appropriate genotype (***complete penetrance***); many genes fail to do so (***incomplete penetrance***).

Peptide Bond. The covalent chemical bond formed due to the reaction of the amino ($—NH_2$) group of one amino acid with the carboxyl (—COOH) group of another amino acid; it links the various amino acids of a polypeptide with each other.

Pericentric Inversion. An inversion in which centromere is located within the inverted segment.

Perinuclear space. The space between inner and outer membranes of nuclear envelope.

Phage. See, ***Bacteriophage.***

Phenocopy. An individual whose phenotype has been modified by the environment to such an extent that it is similar to that produced by another, usually a mutant, allele.

Phenotype. The observable characteristics (or appearance, anatomy etc.) of an organism. (*cf.*, genotype).

Phylogeny. The evolutionary history of a species.

Plaque. The clear area or zone produced by a virus on an opaque lawn of the host bacterium; generally, plated on an agar medium in a petri plate.

Plasmagene. A gene located in a cytoplasmic organelle, *e.g.*, plastid, mitochondria; responsible for cytoplasmic inheritance.

Plasmid. In bacteria; a DNA molecule other than the bacterial chromosome that is capable of independent replication and transmission; it may exist independent of or become integrated into the bacterial chromosome; generally, not essential, except under specific environments.

Plastid. In green plants and some protozoa; an organelle surrounded by two concentric membranes, has DNA, and usually differentiates into typical chloroplasts when exposed to light.

Pleiotropy. A single gene affecting more than one character. *Adj.*, pleiotropic.

Point Mutation. Mutation in a genes due to a change in a single base pair.

Polar Mutation. A mutation, which affects the expression of other genes located downstream of the mutant gene but in the same operon of prokaryotes; due to the formation of a nonsense codon.

Polyadenylation. Addition of poly(A) sequence to the 3′-end of eukaryotic mRNAs after their transcription.

Polycistromic mRNA. The mRNA that codes for 2 or more proteins.

Polygenes. Genes having individually small but cumulative effect on a character; they govern quantitative characters.

Polymorphism. Two or (usually) more phenotypes maintained in a random mating population; due to the effects of environment.

Polynucleotide. A macromolecule in which several nucleotides are joined by the phosphodiester linkage, *e.g.*, DNA and RNA.

Polypeptide. A macromolecule in which several amino acids are joined by the peptide bond.

Polyploid. A cell or an organism having more than two genomes; they may be copies of a single genome (***autopolyploid***) or of different genomes (***allopolyploid***).

Polyprotein. A protein that is cleaved into two or more functional proteins.

Polysome. Polyribosome; an mRNA associated with several ribosomes that are translating the mRNA.

Polytene Chromosome. A giant chromosome made up of several (upto 1,024 or more) copies of itself produced by endomitosis; usually, the homologues are paired as if during pachytene; in *Drosophila,* characteristic bands present. *Syn.*, giant chromosome, salivary gland chromosome.

Population. In ***population genetics;*** a group of interbreeding individuals sharing the same gene pool. In ***statistics;*** all the possible individuals or observations of one kind from which samples are drawn for various studies.

Population Size, Effective. The number of sexually reproducing individuals in a population.

Population Genetics. The branch of genetics concerned with the study of frequencies of genes and genotypes in random mating populations.

Position Effect. Alteration in the phenotypic expression of a gene due to a change in its position in the chromosome; usually, when a gene is placed within or close to a heterochromatin segment, the gene is not expressed due to heterochromatinization.

Primary Transcript. The original unmodified RNA that corresponds to a transcription unit.

Primer. An oligonucleotide (usually, RNA in DNA replication; DNA in PCR) that pairs with one strand of the template DNA, and provides a free 3′-OH essential for DNA polymerase to extend the primer through DNA synthesis.

Prion. A proteinaceous infectious agent, *e.g.*, PrP^{sc}, the causal agent of scrapie in sheep.

Probability. The chance or likelihood of the occurrence of an event.

Processed Pseudogene. A pseudogene that lacks introns; originated by reverse transcription of mRNA and insertion of a duplex copy into the genome.

Prokaryote. Organisms without true nuclei, *i.e.*, the chromosomes are not enclosed by a typical nuclear envelope; they do not undergo meiosis, *e.g.*, viruses, bacteria, cyanobacteria.

Promoter. The DNA segment of an operon or a transcription unit, to which RNA polymerase binds before it initiates transcription.

Proof-reading. In DNA replication, correction of errors that involves scrutiny of nucleotides after they are added to the polynucleotide chain, and their replacement by the correct nucleotides.

Prophase. The first stage of cell division after interphase and before metaphase; the chromosomes become shorter and thicker, and are visibly double-stranded toward the end (mitosis, prophase II of meiosis); chromosomes pair to form typical bivalents (prophase I of meiosis).

Prototroph. A microorganism, which is able to grow normally on a minimal medium containing inorganic salts and a carbon source.

Proto-oncogene. A normal cellular gene, which may change into an oncogene, most likely due to mutation.

Provirus. A duplex DNA sequence in an eukaryotic genome corresponding to the genome of a retrovirus.

Pseudoalleles. Two alleles each of two or more tightly-linked genes affecting the same function as a result of which they appear as multiple alleles; once an

attractive idea for explaining many cases of multiple alleles, *e.g.*, of Rh blood groups; *the hypothesis is not widely accepted.*

Pseudogene. An inactive gene similar to a normal gene and occurs as a normal component of an eukaryotic genome; may arise due to mutation following duplication of the concerned gene.

Pureline. The progeny of a single, homozygous, self-pollinated plant.

Qualitative Character. A character showing a limited number of distinct classes, governed by one or few oligogenes, and relatively little affected by the environment.

Quantitative Character. A character showing continuous variation, governed by polygones, and markedly affected by the environment.

Quantitative Inheritance. The inheritance of quantitative traits due to polygenes.

Random Genetic Drift. The fluctuation in gene frequencies in small random mating populations due to sampling error.

Random Mating. A system of mating in which a kind of male gamete has the same opportunity of uniting with any kind of female gamete of the population; or an individual of a genotype in the population has equal probability of mating with an individual of any genotype present in the population.

Random Mating Population. A population in which random mating is the normal mode of reproduction.

Reading Frame. One of the three possible ways of reading a nucleotide sequence as a series of triplets.

Receptor. A transmembrane protein located in plasma membrane that binds specifically to a small molecule (called ligand) in a domain located outside the cell; the activity of its cytoplasmic domain changes as a result thereof.

Recessive. An allele, which is incapable of expressing itself phenotypically in the heterozygous state.

Reciprocal Crosses. The two crosses in which the same two parents are involved, but the strain that serves as the male parent in one cross is used as the female parent in the other, and *vice-verse.*

Recombination. Usually, in the case of linked genes or characters; the production of new combinations of genes or characters not present in the parents of a cross or generation; involves crossing over.

Recon. The unit of recombination; is represented by one nucleotide; the part of gene within which recombination cannot take place.

Reduction Division. The first division of meiosis when the two chromosomes forming each bivalent separate and move to the opposite poles.

Regulator Gene. The gene of an operon, which codes for the repressor or activator of the operon, *e.g.*, the gene *i* of *E. coli lac* operon.

Renaturation. The reassociation of complementary strands of DNA to form duplexes; occurs after denaturation.

Repetitive DNA. DNA sequences present in many (two or more) copies (related or identical) per genome.

Replication. The process of production of new copies of a macromolecule by exactly copying the preexisting macromolecule, *e.g.*, DNA ⟶ DNA or RNA ⟶ RNA.

Replication Fork. During DNA replication; the points at which the two strands of the parental DNA duplex are separated to allow replication to proceed.

Replicon. A segment of DNA capable of undergoing replication; it has a unique 'origin' site at which the replication begins.

Replisome. The complex of proteins that forms at the bacterial replication forks; contains DNA polymerase.

Reporter Gene. A gene whose product is easily assayed; ***scorable*** (product assayble) or ***selectable*** (product permits an easy selection of the cells expressing it) reporters.

Repressible Enzyme. An enzyme that is produced by the cell as long as its end-product does not accumulate to a threshold concentration.

Repression. Prevention of transcription of genes in response to accumulation of the products of enzymes encoded by these genes.

Repressor Protien. This protein binds to operator DNA to generate repression.

Repulsion Phase. A linkage between the dominant allele(s) of one (or more) gene(s) and the recessive allele(s) of another (several other) gene(s); in such a case one parent involved in a cross contributes the dominant allele of one (or more) gene(s), while the second parent provides the dominal allele(s) of the other gene(s).

Resistance Factor. A plasmid, which carries genes that confer to its host bacterium resistance to one or more antibiotics.

Restitution. When a chromosome breaks and the two broken ends thus produced join in their original position to yield the original chromosome.

Restitution Nucleus. A nucleus, which contains all the chromosomes present in a cell following the failure of a cell to undergo complete division.

Restriction Fragment Length Polymorphism (RFLP). Differences in the lengths of fragments produced by cleavage with the relevant restriction enzyme of the same stretch of DNA from different individuals/strains/species; due to heritable differences in the locations of recognition site for the enzyme.

Restriction Enzyme. An endonuclease, which cleaves a DNA molecule either within or close to a site having a specific base sequence.

Restriction Map. A map depicting the target sites for various restriction enzymes on a DNA molecule.

Retroposon. A transposon that mobilizes *via* an RNA form; DNA form → RNA form → DNA form (integrates into host genome).

Retrovirus. An RNA virus that propagates *via* a DNA duplex form.

Reunion. The joining of broken chromosome ends, following chromosome breakage, in ways other than they were originally united.

Reverse Mutation. The mutation of a mutant allele back to the wild type allele.

Reverse Transcription. Synthesis of DNA using RNA as a template; catalyzed by reverse transcriptase.

Rh Factor. An antigen present on the RBCs of some human beings, which is antigenically similar to an antigen found in *Macaca rhesus* monkeys.

Rho Factor. In *E. coli.*, the protein that helps RNA ploymerase to terminate transcription at certain sites (called ***rho-dependent terminators***).

Ribonucleic Acid. A polynucleotide having ribose as the pentose sugar and ordinarily uracil in place of thymine; genetic material in many viruses and may undergo replication; generally, single-stranded; is of five types: mRNA, tRNA, rRNA, small nuclear or cytoplasmic RNAs, and genetic RNA; usually, produced by the transcription of DNA.

RNAase. An enzyme that degrades RNA.

RNA Polymerase. The enzyme that catalyzes transcription, *Syn.*, dNA-dependent RNA polyewrase.

RNA Replicase. The enzyme that synthesizes RNA using RNA as template; catalyzes RNA replication in certain virsus.

S Phase. In eukaryotes; the stage between G1 and G2 when DNA replication takes place.

Saltatory Replication. A sudden lateral amplification to produce a large number of copies of some DNA sequence.

Satellite DNA. In genomic DNA; a narrow peak on a density gradient, representing many tandem repeats of related or identical short sequences.

Segregation. Separation of the two alleles of a gene during gamete formation and their passage into different gametes; usually, due to the separation of homologous chromosomes during A I; each gamete contains only a single copy of only one allele, and leads to the typical 3 : 1 or 1 : 2 : 1 ratios in the F_2 generation of monohybrid crosses.

Segregation Distorter. A gene or some other genetic element, which alters the ratio of segregating alleles recovered in the gametes; in *Drosophila,* the SD^+ cells are eliminated during gametogenesis.

Selection. Differential reproduction rates of the different genotypes in a population; the most important factor, which determines the direction and the rate of evolutionary change in a population.

Self-fertilization. The union between male gametes of an individual with the female gametes produced by the same individual; usually in plants, ***self-pollination.***

Self-incompatibility. The inability of fully functional pollen grains to effect fertilization and seed set on self-pollination; usually, due to multiple alleles of a single gene *S;* gametophytic or sporophytic in nature.

Semiconservative Replication. DNA replication; progressive separation of the two parental duplexes, each of which serves as template for the synthesis of a new complementary strand to yield two daughter duplexes; each daughter duplex has one old and one newly synthesized strand.

Sense Strand. The strand of DNA duplex that has the same base sequence as mRNA.

Sex Chromosome. A chromosome, which differs in number or kind between the male and female individuals of the same species, and is involved in sex determination. *Syn.,* allosome.

Sexduction. In bacteria; the incorporation of bacterial genes into a sex factor, *e.g., F,* and their subsequent transfer to other bacterial cells through conjugation.

Sex Factor. See, ***F Factor.***

Sex-influenced Trait. Alleles of the gene governing such a trait show opposite dominance relationship in the males and females of the concerned species.

Sex-limited Character. A character, which is expressed in one of the two sexes only.

Sex-linkage. The association between a character and the sex during inheritance; the concerned gene is located in the *X* chromosome; in human beings, several genetic diseases are sex-linked.

Sex Mosaic. See, ***Gynandromorph.***

Sexual Reproduction. Production of new individuals through the fusion of male and female gametes.

Shine-Dalgarno Sequence. In bacterial mRNAs; sequence AGGAGG located upstream of the AUG initiation codon; complementary to a sequence at the 3′-end of 16*S* rRNA; involved in binding of ribosomes to mRNA.

Shot-gun Cloning. All the fragments generated from genomic DNA used for cloning.

Sib-mating. Mating between individuals, which are progeny of the same two parents. *Syn.*, brother-sister mating.

Sigma Factor. In bacteria; the subunit of RNA polymerase needed for transcription initiation specifically at the promoters; not needed for the subsequent transcription.

Signal Sequence. In proteins; the sequence, usually N-terminal, needed for insertion of the protein into the membranes of ER concomitant with its translation.

Simple Sequence DNA. See, ***Satellite DNA.***

Site-specific Recombination. It occurs between two specific, but not necessarily homologous, sequences; *e.g.,* during λ phage integration/excision.

snRNA. Small nuclear RNA; a species of small RNAs that are confined to the nucleus; involved in mRNA splicing and other RNA Precessing reactions.

Somatic Cell. A cell making up the body of an organism, excluding the germ cell; such cells have the full somatic complement of the species.

Somatic Chromosome Complement. The number and the type of chromosomes normally found in the somatic cells, particularly in the meristematic cells, of a species; it is twice the gametic complement (*n*) and is denoted by 2*n*.

Somatic Chromosome Number. The chromosome number normally found in the somatic cells, more specifically in the meristematic cells, of a species; represented by 2*n*.

Somatic Mutation. A mutation in a somatic cell, usually in meristematic cell. *Syn.,* bud mutation.

Southern Blotting. Procedure for transfer of denatured DNA from an agarose gel onto a nitrocellulose filter to allow its hybridization with a complementary nucleic acid (called ***probe***).

Species. A group of interbreeding or potentially interbreeding individuals, which is reproductively isolated from other such groups.

Sperm. A mature male gamete; *abbreviation* of spermatozoon; *pl.,* spermatozoa.

Spermatogenesis. The process of production of mature male gametes or sperms from diploid sperm mother cells through meiosis; in case of animals.

Splicing. In eukaryotes; removal of introns and joining of the exons (in their natural order) to generate mature mRNAs.

Spontaneous Mutation. A mutation that arises naturally without the application of any mutagen (by man); the cause of such mutations is not definitely known.

Sporogenesis. The production of haploid micro- and megaspores from the diploid microspore and megaspore mother cells, respectively, through meiosis; in plants.

Sporulation. Production of spores by a bacterium or by yeast.

Staggered Cuts. The two strands of a DNA are cleaved at different points near each other; generate protruding ends.

Standard Deviation. The Square root of variance; a measure of the amount of variation present in a population.

Standard Error. A measure of variability in a population of sample means derived from a single population; an indicator of the correspondence between sample estimates and population parameters; estimated from a sample as $\sqrt{\text{Variance} / n}$ (where n is the number of observations in the sample).

Startpoint. The site on a DNA molecule that corresponds to first base incorporated into RNA during transcription. *Syn.*, startsite.

Statistic. An estimate of a population parameter, *e.g.*, population mean, obtained from a sample drawn from the concerned population.

Sterility. Inability to produce progeny usually due to nonfunctional gametes; male and female sterility.

Structural Gene. A gene, which codes for (through mRNA) the amino acid sequence of a polypeptide.

Sublethal Gene. A gene, which causes the death of most, but not all, of the individuals that carry it in the appropriate genotype.

Subvital Gene. A gene, which kills some of the individuals that carry it in the appropriate genotype.

Superfemale. See, ***Metafemale.***

Supermale. A weak and almost completely sterile male individual exhibiting more intense secondary sex characters than normal males; in *Drosophila,* due to an excess of autosomal sets in relation to the number of X chromosomes as compared to that in the normal males; $X/A = 0.5$ = male; $X/A < 0.5$ = supermale.

Suppressor Mutation. A mutation, which suppresses the phenotypic expression of another mutation; usually, in another gene, or in the same gene at a different site.

Symbiont. An organism living in a close association with another organism in such a way that both of them benefit from the association.

Synapsis. During prophase I of meiosis; close, longitudinal, point-to-point precise association between homologous chromosomes. *Syn.*, chromosome pairing.

Synaptonemal Complex. A proteinaceous structure about 1000 Å wide formed between homologous chromosomes during synapsis; suggested to facilitate recombination or crossing over.

Syndrome. A group of symptoms, which occur together and characterise a particular disease.

Synteny. Two or more genes being located in the same chromosome without any reference to their linkage relationship. *Adjective,* syntenic.

TATA Box. In eukaryotic promoters; a conserved A•T-rich 7 base sequence located ~25 bp upstream of the startpoint of all genes transcribed by RNA polymerase II; may be involved in the correct positioning of RNA polymerase II.

Telocentric. A chromosome having centromere at one end (in the place of telomere).

Telomerase. A ribonucleoprotein enzyme that adds repeating units of telomere to one DNA strand; individual bases added to the 3′-end of the strand.

Telomere. The natural end of a chromosome; consists of a simple repeating sequence, and a single protruding end that may fold back to form a hairpin.

Telophase. The last stage of cell division following anaphase preceding interphase; the chromosomes gather at the opposite poles and begin to uncoil to finally form loosely packed masses of chromatin fibers.

Temperate Phage. A phage, which may integrate into the chromosome of its host and may remain there as a prophage (***lysogeny***), or may dissociate from the chromosome of its host and lyse it like a virulent phage.

Temperature-Sensitive Mutation. The product of mutant gene is functional at low temperature, but inactive at high temperature. (The reverse situation is called ***Cold-Sensitive***).

Template. A mould; a substance used to create identical copies as specified by itself; DNA molecules serve as template during replication and transcription.

Terminalization. More commonly, ***chiasma terminalization;*** the movement of visible chiasmata toward the ends of chromosome arms due to the movement away from each other of the homologous chromosomes in each bivalent; the movement is more pronounced at the centromeres; terminalization begins in the diplotene of prophase I of meiosis and ends at the onset of diakinesis.

Termination Codon. See ***Nonsense Codon.***

Terminator. The DNA sequence that causes RNA polymerase to terminate transcription; located at the end of each transcription unit.

Testcross. The cross of an F_1 hybrid with an individual or strain having the recessive phenotype for the concerned trait.

Tetrad. The group of four daughter cells produced by meiotic division of a single cell; also, the four-stranded structure (*i.e.*, bivalent) produced due to the synapsis of two homologous chromosomes, each having two chromatids.

Tetraploid. A cell or an organism having four genomes; the genomes may be copies of a single genome (***autotetraploid***) or of different genomes (***allotetraploid***).

Tetrasomic. A cell or an organism having one chromosome pair in addition to the normal somatic complement of the species; one chromosome of the genome has four copies or homologoues.

Threshold. The concentration, level or intensity below which a physical or chemical agent does not produce its effect.

Topoisomerase. An enzyme that can induce or relax supercoils in DNA, *i.e.*, change the linking number (in steps of 1) of a DNA molecule.

Totipotent Cell. A cell capable of undergoing development like the zygote, that is, of developing into a normal embryo, under a suitable environment.

Trailer. The nontranslated sequence at the 3′-end of an mRNA molecule; downstream of the termination codon.

Trans-*Arrangement.* See, ***Repulsion Phase.***

Transcription. The production of an RNA molecule complementary to a DNA molecule, which serves as the template for RNA polymerase, the enzyme that catalyzes transcription.

Transcription Unit. The stretch of DNA located between a startpoint (*site of transcription initation*) and terminator (*site of transcription termination*).

Transduction. Genetic recombination in bacteria following the transfer by a virus of a segment of the bacterial chromosome from one bacterial cell into another.

Transduction, Abortive. The bacterial DNA transferred by a virus does not undergo recombination, fails to replicate and is gradually diluted and lost from the progeny cells.

Transduction, Generalized. The phage may transfer any bacterial gene to the recipient cell; a phage particle contains a random segment of the bacterial chromosome; mediated by virulent phages.

Transduction, Restricted. The phage transfers only those genes that are located near a specific site, *i.e.,* the site where a temperate phage integrates into the bacterial chromosome; mediated by temperate phages.

Transfer RNA (tRNA). RNA molecules that transport specific amino acids to the ribosome; each tRNA species has a specific anticodon, which base-pairs with the appropriate mRNA codon.

Transfection. The uptake of DNA by eukaryotic cells, followed by the incorporation of genes present in this DNA into the genome of eukaryotic cells.

Transformation. In bacteria, genetic recombination following the uptake of a DNA molecule by a competent bacterial cell; in animals, the conversion of a cell growing *in vitro* to a state of uncontrolled growth similar to that of a tumour cell; in genetic engineering, introduction of a vector containing the desired gene into a host cell.

Transgene. A gene introduced into an organism by genetic engineering; usually, the gene is expressed.

Transgenic. An organism that contains in its genome and (usually) expresses a transgene.

Transgressive Segregation. The appearance of individuals in the F_2 or a subsequent generation, which exceed the parental limits with respect to one or more characters.

Transition. The substitution, in a DNA or an RNA molecule, of a purine by another purine or of a pyrimidine by another pyrimidine; leads to gene mutation.

Translation. Synthesis of a polypeptide in association with a ribosome and directed by an mRNA molecule with the help of tRNA and several highly specific enzymes.

Translocation. Integration of a segment of one chromosome into a nonhomologous chromosome; this integration may be reciprocal, *i.e.,* the two nonhomologous chromosomes involved in the translocation may exchange segments; *in case of a gene:* appearance of a new copy of the gene at a new location in the genome; *in case of a protein:* movement across a membrane.

Transmembrane Protein. A protein that is component of a membrane; its hydrophobic domain or domains reside in the membrane, while its hydrophilic regions are exposed on one or both the sides of the membrane.

Transposase. The enzyme that sponsors movement of transposons to new sites in the genome.

Transposon. A DNA segment, which is capable of changing its location within a chromosome or between chromosomes.

Transposition. Movement of a tranposable element to a new site in the genome; sponsored by transposase.

Transversion. In DNA or RNA; substitution of a purine by a pyrimidine or *vice-versa.*

Trihybrid. The F_1, F_2 etc. progeny from a cross involving two parents differing for three genes (or characters), or considered to be so for that study.

Trisomic. A cell or an organism having one chromosome in excess of the normal somatic complement of the species; one chromosome has three homologues in place of the normal two; denoted by $2n + 1$.

Turner's Syndrome. In human beings; individuals having XO chromosome constitution, being phenotypically female but having rudimentary sexual organs and mammary glands.

Unequal Crossing Over. The sites of recombination are located at different sites in the two homologous DNA duplexes.

Univalent. An unpaired chromosome during prophase I and metaphase I of meiosis.

Upstream. The sequences proceeding in the opposite direction from the direction of expression; in case of a transcription unit, sequences lying on the left of the startpoint.

URF. Unidentified (open) reading frame; presumed to encode protein, but no protein product has been found.

Variance. A measure of variability; it equals the sum of the squares of deviations of all the observations in a sample from the sample mean divided by one less than the number of observations; represented as, $[\Sigma(X - \bar{X})^2]/(N - 1)$.

Variation. The occurrence of differences among the individuals belonging to a single species for one or more traits.

Vector. A plasmid or phage that is used to 'carry' inserted foreign DNA for the purposes of producing more copies of the DNA insert or a protein product.

Virulent Phage. A virus or phage, which always causes the lysis of its host bacterium following infection.

Wild Type. The phenotype common in the wild population of a species or found normally in a species.

Wobble Hypothesis. A hypothesis to account for the degeneracy of genetic code; the first two bases of a codon and its tRNA anticodon pair properly or specifically, while the pairing at the third base has some play or wobble so that the same anticodon may pair with two or more codons differing at the third base.

X Chromosome. The sex chromosome, which occurs in both male and female individuals but in different numbers, *e.g.,* human male has one *X* chromosome, while females have two; in most animals and many plants, it has the genes for femaleness.

Y Chromosome. The sex chromosome, which occurs either in the males (man, rodents, *Drosophila* etc.) or females (birds, reptiles) of a species, but never in both the sexes; usually the major portion of *Y* is genetically inactive; may or may not determine maleness depending on the species.

Transcription Unit. The stretch of DNA located between a startpoint (*site of transcription initation*) and terminator (*site of transcription termination*).

Transduction. Genetic recombination in bacteria following the transfer by a virus of a segment of the bacterial chromosome from one bacterial cell into another.

Transduction, Abortive. The bacterial DNA transferred by a virus does not undergo recombination, fails to replicate and is gradually diluted and lost from the progeny cells.

Transduction, Generalized. The phage may transfer any bacterial gene to the recipient cell; a phage particle contains a random segment of the bacterial chromosome; mediated by virulent phages.

Transduction, Restricted. The phage transfers only those genes that are located near a specific site, *i.e.,* the site where a temperate phage integrates into the bacterial chromosome; mediated by temperate phages.

Transfer RNA (tRNA). RNA molecules that transport specific amino acids to the ribosome; each tRNA species has a specific anticodon, which base-pairs with the appropriate mRNA codon.

Transfection. The uptake of DNA by eukaryotic cells, followed by the incorporation of genes present in this DNA into the genome of eukaryotic cells.

Transformation. In bacteria, genetic recombination following the uptake of a DNA molecule by a competent bacterial cell; in animals, the conversion of a cell growing *in vitro* to a state of uncontrolled growth similar to that of a tumour cell; in genetic engineering, introduction of a vector containing the desired gene into a host cell.

Transgene. A gene introduced into an organism by genetic engineering; usually, the gene is expressed.

Transgenic. An organism that contains in its genome and (usually) expresses a transgene.

Transgressive Segregation. The appearance of individuals in the F_2 or a subsequent generation, which exceed the parental limits with respect to one or more characters.

Transition. The substitution, in a DNA or an RNA molecule, of a purine by another purine or of a pyrimidine by another pyrimidine; leads to gene mutation.

Translation. Synthesis of a polypeptide in association with a ribosome and directed by an mRNA molecule with the help of tRNA and several highly specific enzymes.

Translocation. Integration of a segment of one chromosome into a nonhomologous chromosome; this integration may be reciprocal, *i.e.,* the two nonhomologous chromosomes involved in the translocation may exchange segments; *in case of a gene:* appearance of a new copy of the gene at a new location in the genome; *in case of a protein:* movement across a membrane.

Transmembrane Protein. A protein that is component of a membrane; its hydrophobic domain or domains reside in the membrane, while its hydrophilic regions are exposed on one or both the sides of the membrane.

Transposase. The enzyme that sponsors movement of transposons to new sites in the genome.

Transposon. A DNA segment, which is capable of changing its location within a chromosome or between chromosomes.

Transposition. Movement of a tranposable element to a new site in the genome; sponsored by transposase.

Transversion. In DNA or RNA; substitution of a purine by a pyrimidine or *vice-versa.*

Trihybrid. The F_1, F_2 etc. progeny from a cross involving two parents differing for three genes (or characters), or considered to be so for that study.

Trisomic. A cell or an organism having one chromosome in excess of the normal somatic complement of the species; one chromosome has three homologues in place of the normal two; denoted by $2n + 1$.

Turner's Syndrome. In human beings; individuals having *XO* chromosome constitution, being phenotypically female but having rudimentary sexual organs and mammary glands.

Unequal Crossing Over. The sites of recombination are located at different sites in the two homologous DNA duplexes.

Univalent. An unpaired chromosome during prophase I and metaphase I of meiosis.

Upstream. The sequences proceeding in the opposite direction from the direction of expression; in case of a transcription unit, sequences lying on the left of the startpoint.

URF. Unidentified (open) reading frame; presumed to encode protein, but no protein product has been found.

Variance. A measure of variability; it equals the sum of the squares of deviations of all the observations in a sample from the sample mean divided by one less than the number of observations; represented as, $[\Sigma(X - \bar{X})^2]/(N - 1)$.

Variation. The occurrence of differences among the individuals belonging to a single species for one or more traits.

Vector. A plasmid or phage that is used to 'carry' inserted foreign DNA for the purposes of producing more copies of the DNA insert or a protein product.

Virulent Phage. A virus or phage, which always causes the lysis of its host bacterium following infection.

Wild Type. The phenotype common in the wild population of a species or found normally in a species.

Wobble Hypothesis. A hypothesis to account for the degeneracy of genetic code; the first two bases of a codon and its tRNA anticodon pair properly or specifically, while the pairing at the third base has some play or wobble so that the same anticodon may pair with two or more codons differing at the third base.

X Chromosome. The sex chromosome, which occurs in both male and female individuals but in different numbers, *e.g.,* human male has one *X* chromosome, while females have two; in most animals and many plants, it has the genes for femaleness.

Y Chromosome. The sex chromosome, which occurs either in the males (man, rodents, *Drosophila* etc.) or females (birds, reptiles) of a species, but never in both the sexes; usually the major portion of *Y* is genetically inactive; may or may not determine maleness depending on the species.

Zoo Blot. Use of Southern blotting to test the ability of a DNA probe from one species to hybridize with the DNA from genomes of a variety of other species.

Zygote. The cell derived through fusion of one male and one female gamete; it has $2n$ chromosome complement as opposed to the n for the gametes; it develops into a new individual through mitosis (higher plants and animals) or may undergo meiosis to yield the haploid phase (*e.g.,* fungi).

Zygotene. A substage of prophase I of meiosis following leptotene and before pachytene; homologous chromosomes pair during this stage to form bivalents.

Appendix-I

A CHRONOLOGY OF IMPORTANT DEVELOPMENTS RELATED TO GENETICS

Year	*Scientist*	*Contribution/Discovery*
700 B.C.	—	Babylonians and Assyrians pollinated date palm to ensure good fruit set.
400 B.C.	Hippocrates	Suggested inheritance of acquired characters.
1590	Z. Janssen and H. Janssen	Assembled the first compound microscope.
1651	W. Harvey	Proposed that all living things originate from eggs.
1657	R. de Graaf	Follicles in human ovary (thought them to be eggs); *suggested that both parents contributed to heredity.*
1665	R. Hooke	First description of cells (dead cells); published *Micrographia.*
1668	R. Redi	Disproves spontaneous generation of fly larvale.
1677	A. Van Leeuwenhoek	Observes sperms of mammals, including man
1694	J. R. Camerarius	Pollination essential for seed set in maize.
1717	Thomas Fairchild	Produced first artificial plant hybrid: carnation × Sweet William (= Fairchild's Mule).
1761-67	J. G. Koelreuter	Extensive hybridization between *Nicotiana* spp.; ***observed***: F_1's are uniform and intermediate between the parents; F_2's much more variable; reciprocal F_1's comparable; ***concluded*** that both the parents contribute equally to the characteristics of offspring.
1769	L. Spallanzani	Disproves spontaneous generation of microorganisms.
1780	L. Spallanzani	Artificial insemination in amphibians and later (1785) in dogs.
1798	T. R. Malthus	Publishes anonymously 'An Essay on the Principle of Population', which later suggested the concept of struggle for existence to Darwin.

1809	J. B. de Monet Lamarck	Proposes evolution of new species through inheritance of acquired characters.
1818	W. C. Wells	Puts forth the idea of natural selection.
1820	C. F. Nasse	Describes sex-linked pattern of inheritance of hemophilia in man.
1822-24	T. A. Knight, J. Goss and A. Seton	Observed dominance in F_1 and segregation in F_2, but did not propose any theories of inheritance.
1827	K. E. von Baer	First accurate description of human egg.
1830	G. B. Amici	Pollen tube grows into style and ovule.
1831	R. Brown	Notes presence of nuclei in cells.
1838-39	M. J. Schleiden and T. Schwann	*Propose cell theory*; also Schleiden notes nucleoli within nuclei.
1845	J. Dzierzon	In honey bees, drones hatch from unfertilized eggs, while workers and queens hatch from fertilized eggs.
1855	A. R. Wallace	Begins to develop theory of evolution similar to that of Darwin.
	R. Virchow	Proposed cell lineage theory.
1856	G. J. Mendel	Begins studies on hybridization in peas.
1858	C. Darwin and A. R. Wallace	Jointly present a paper to the Linnean Society of London containing the theory of evolution through natural selection.
1859	Charles Darwin	Publishes the book '*On the Origin of Species*'.
1861	L. Pasteur	Conclusively disproves spontaneous generation of microorganisms.
1865	G. J. Mendel	Presented the historical paper 'Experiments in Plant Hybridization' to the "Brünn Society for the Study of Natural Sciences"; the paper was published in the proceedings of the Society in 1866.
1866	L. Down	Describes Down's syndrome
1869	F. Galton	Describes a scientific study of human pedigree in his book '*Hereditary Genius*'.
1871	F. Meischer	Isolated nucleic acids; called them *nuclein.*
1873	A. Schneider	First account of mitosis.
1875	F. Galton	Studied twins to elucidate the contributions of nature (heredity) and nurture (environment) to human behaviour.
	O. Hertwig	Concludes that fertilization in animals and plants involves a physical fusion of egg and sperm nuclei.

	E. Strausberger	Describes cell division in plants.
1879	W. Flemming	Demonstrates longitudinal division of chromosomes during mitosis; coins '*chromatin*'.
1881	E. G. Balbiani	Discovers giant chromosomes in salivary glands of *Chironomus* but fails to realize they are chromosomes.
1882	W. Flemming	Discovers lampbrush chromosomes; coins '*mitosis*'.
1883	E. Van Beneden	Studies meiosis in *Ascaris*; gametes have only half as many chromosomes as somatic cells; fertilization restores chromosomes number to somatic number.
	A. Weismann	Proposes germplasm theory.
1884	E. Strausberger	Fusion of egg and sperm nuclei in angiosperms.
1888	T. Boveri	Describes contriole.
	E. Strausberger	Reduction in chromosome number during gametogenesis.
	W. Waldeyer	Coins the term '*chromosome*'.
1889	F. Galton	Lays the foundations of 'biometry' in his book '*Natural Inheritance*'.
1890	T. Boveri	Integrity of metaphase chromosomes maintained during interphase.
1896	E. Wilson	Publishes '*The Cell in Development and Heredity*'.
1899	M. W. Beijerinck	Demonstrates the tobacco mosaic to be due to a self-replicating agent, which he called '*virus*'
1900	Hugo de Vries, C. Correns and E. Tschermak	Rediscover Mendel's paper; they had independently arrived at conclusions similar to those of Mendel.
	W. Bateson	Stresses the importance of Mendel's finding in an addess to Royal Society of London.
	K. Pearson	Develops chi-square test.
1901	H. de Vries	Coins the term '*mutation*'.
	K. Landsteiner	Human ABO blood groups.
1902	C. E. McClung	Discovers 'accessory chromosome' (= X chromosome) in insects; suggest that it determines sex.
	T. Boveri	Studies development of haploid, diploid and aneuploid sea urchin embryos; concludes that individual chromosomes carry different hereditary determinants (= genes).
	W. Sutton	Proposes 'chromosome theory of heredity'.

1902 - 1909	W. Bateson	Introduces the terms : *genetics, allelomorph, homozygote, heterozygote*, F_1, F_2 and *epistatic genes.*
1905	L. Cuenot	Describes 2 : 1 F_2 ratio in case of yellow mice; concludes : yellow mice are heterozygous.
	Farmer	Coined the term '*meiosis*'.
	Wilson and Stevens	Coined the term '*X chromosome*'
1906	W. Bateson and R.C. Punnet	Report first case of linkage in sweet pea, but misinterpret the data.
	Yule	Suggested the basic idea of multiple factor hypothesis.
1907	E.F. Smith	*Agrobacterium tumefaciens* causes crown gall disease.
1908	G. H. Hardy and W. Weinberg	Independently formulate the Hardy-Weinberg law.
	Wilson an Stevens	Described Y chromosome in *Drosophila.*
1909	A.E. Garrod	Publishe '*Inborm Errors of Metabolism*'; persons suffering from a hereditary disease lacked an enzyme.
	Castle and Phillips	Present conclusive proof of germplasm theory by ovary transplant in guinea pigs.
	W. Johannsen	Proposes 'pureline theory'; coins the terms: *gene, genotype* and *phenotype.*
	C. Correns and E. Baur	Report the first case of 'plastid inheritance (= cytoplasmic inheritance)
	H. Nilsson-Ehle	Proposes 'multiple factor hypothesis'.
	E. Wilson	Gave an accurate account of the role of *X* chromosome in sex determination.
1909 (1924)	Janssens	'Chiasmatype' or 'one-plane theory' of chiasma formation and recombination.
1910	T. H. Morgon	Sex-linkage in *Drosophila.*
1911	T. H. Morgan	Proposes that genes for white eyes, yellow body and miniature wings in *Drosophila* are linked together in the *X* chromosome; describes essential features of linkage.
1912	T.H. Morgan	Crossing over does not take place in male *Drosophila.*
1913	A. H. Sturtevant	Experimental basis for linkage concept; the first genetic map
	Davenport	Explained the inheritance of skin colour in white × negro marriages.
1914	C. B. Bridges	Nondisjunction of *X* chromosome in *Drosophila.*

1916	H. J. Muller	Discovered interference in *Drosophila.*
	C. B. Bridges	The first case of aneuploidy (*XO* males in *Drosophila*).
1917	O. Winge	Points out the role of polyploidy in the evolution of augiosperms.
	C. B. Bridges	First case of deletion (in *Drosophila*).
1918	H. J. Muller	Balanced lethal phenomenon in *Drosophila.*
1919	C. B. Bridges	Discovers duplications in *Drosophila.*
1920	A. F. Blakeslee, J. Belling and M. E. Farnham	Trisomics in *Datura.*
1921	C. B. Bridges	Reports first case of monosomic (haplo-4) in *Drosophila.*
1922	L. V. Morgan	Attached-*X* chromosome in *Drosophila.*
1923	C. B. Bridges	Discovers translocations (in *Drosophila*).
1925	C. B. Bridges	Proposes 'genic balance theory of sex determination'.
	E. M. East and A. J. Mangelsdorf	First satisfactory explanation of self-incompatibility.
	A. H. Sturtevant	Discovers position effect in case of *Bar* eye of *Drosophila.*
	T. H. Goodspeed and R. E. Clausen	Produce the first synthetic amplidiploid (*Nicotiana digluta* from *N. glutinosa X N. tabacum*) in *Nicotiana.*
1926	A. H. Sturtevant	The first case of inversion (in *Drosophila*).
1927	H. J. Muller	Mutation induction by X-rays in *Drosophila.*
1928	L. J. Stadler	Mutation induction in maize by *X*-rays
	F. Griffith	Transformation in *Diplococcus preumoniae.*
	L. F. Randolph	'B. chromosomes' (= accessory chromosomes) in plants.
	E. Heitz	Coins the term '*heterochromatin*'.
1931	C. Stern (independently), H. B. Creighton and B. McClintock	Cytological proof of crossing over.
1933	Rhoades	Cytoplasmic male sterility in maize.
1934	Sharp	The 'classical' or 'two-plane theory' of chiasma formation and crossing over.
1936	C. Stern	Somatic crossing over in *Drosophila.*
1937	T. Dobzhansky	Publishes '*Genetics and the Origin of Species*'
	A. F. Blakeslee and A. G. Avery	Chromosome doubling effect of colchicine.

1938	B. McClintock	Bridge-breakage-fusion-bridge cycle.
	T. M. Sonneborn	Killer factor in *Paramecium.*
	H. H. Turner	Described 'Turners' syndrome'.
	M. M. Roades	Describes the mutator gene *Dt* in maize.
1940	Oliver	Intragenic recombination in the *lz* (*lozenge*) locus of *Drosophila.*
1941	G. W. Beadle and E. L. Tatum	One-gene-one-enzyme theory.
	K. Mather	Coins the term '*polygenes*'.
1942	H. F. Klinefelter	Described 'Kilinefelter's syndrome'.
1944	O. T. Avery, C. M. MacLeod and M. McCratey	Demonstrate that DNA is the transforming principle in *Diplococcus pneumoniae.*
1946	C. Auerbach and J.M. Robson	Mutagenic effects of mustard gas (a chemical).
1950	B. McClintock	*Ac-Ds* system (control elements) in maize.
1952	N. D. Zinder and J. Lederberg	Describe transduction in *Salmonella.*
	A. D. Hershey and M. Chase	DNA is the genetic material in *E. coli* phage T_2.
1953	J. D. Watson and F. H. C. Crick	Proposed the double helix model of DNA.
	C.C. Lindegren	Gene conversion in *Saccharomyces.*
1954	G. Gammow	Proposes the codon to be a triplet.
1955	S. Benzer	Fine structure of the r^{II} locus of T_4 phage of *E. coli*; coins the terms '*cistron*', '*recon*' and '*muton*'
	Volkin and Astrachan	Existence of *m*RNA (studies on T_2-infected *E. coli*)
	H. Frankel-Conrat and R. C. Williams	Demonstrate that RNA is genetic material in TMV.
	F. Jacob and E. L. Wollman	Bacterial conjugation involves transfer of a piece of DNA from donor to recipient cells.
	S. Ochoa and coworkers	*In vitro* synthesis of RNA.
1956	A. Kornberg and coworkers	*In vitro* synthesis of DNA.
	J. H. Tijo and A. Levan	Demonstrate human chromosomes to be 46.
1957	J. H. Taylor, P.S. Woods and W.L. Hughes	Semiconservative replication of *Vicia faba* chromosomes.

	V. M. Ingram	Sickle cell hemoglobin differs from normal hemoglobin by a single amino acid; one-gene-one-polypeptide hypothesis.
1958	F. Jacob and E. L. Wollman	*E. coli* linkage group is circular.
	F.H.C. Crick	Proposed the participation of an adapter (= *t*RNA) in protein synthesis; DNA (albeit via RNA) determines the amino acid sequence of polypeptides.
	P.C. Zamecnik and coworkers	Characterize amino acid-transfer RNA complexes.
	M. Okamoto (independently), R. Riley and V. Chapman	Discover genes (*Ph*) that controls homologous chromosome pairing in wheat.
	M. Meselson and F. W. Stahl	Demonstrate semiconservative replication of DNA in *E. coli.*
1959	J. Lejeune, M. Gautier and R. Turpin	Down syndrome is due to the trisomy for a chromosome (chromosome 21).
	C. E. Ford and coworkers	Turner syndrome females are *XO.*
	P.A. Jacobs and J.A. Strong	Klinefelter males are *XXY.*
	S.B. Wiess	RNA polymerase activety in rat liver.
	A. Lima-de-Faria	Heterochromatin replicates later than euchromatin.
	K. McQuillen, R.B. Roberts and R.J. Britten	*E. coli* ribsomes are sites of protein synthesis.
	R.H. Whittaker	Classifies organisms into 5 kingdomes: (1) bacteria, (2) eukaryotic microorganisms, (3) fungi, (4) plants, and (5) animals.
1961	F. Jacob and J. Monod	Propose the operon concept of gene regulation; proposed the existence and function of *m*RNA.
	F. Jacob, M.F. Lyon and several others	Independent demonstration of inactivation of one *X* chromosome in human females.
	S.B. Weiss and T. Nakamoto	Isolate RNA polymerase.
	G. von Ehrenstein and F. Lipman	Cell-Free system for protein synthesis.

	F.H.C. Crick, L. Barnett, S. Brenner and R.J. Watts - Tobin	Genetic code has 3-letter words (= three bases together code for one amino acid); genetic code is nonoverlapping, commaless and degenerate.
	M.W. Nirenberg and J.H. Matthaei	The meaning of first triplet codon: UUU encodes phenylalanine; TMV RNA or poly-URNA supports polypeptide synthesis *in virto*.
	M. Meselson and J.J. Weigle	Recombination in phage λ involves chromosome breakage and reunion, but not replication.
1962	H. Ris and W. plaut	Chloroplasts contain DNA (by electron microscopy).
	J.B. Gurdon	Obtained normal fertile frog from an enucleated egg injected with a nucleus from an intestine cell.
1963	J. Monod and S. Brenner	The replicon model.
	J. Cairns	Autoradiographic picture of replicating chromosome of *E. coli*; circular, with *Y*-shaped replication forks.
	E. Mayr	Publishes '*Animal Species and Evolution*'.
	Whitehouse	'Hybrid DNA Model of recombination'; breaks in strands of opposite polarity.
1964	R. Holiday	'Hybrid DNA Model or recombination; breaks in strands of same polarity.
	R.B. Setlow and W.J. Carrier, and (independently) R.P. Boyce and P. Howard - Flandess	Excision-repair in bacteria.
	A.S. Sarbhai and coworkers (independently), C. Yanofski and coworkers	Colinearity of genes and their protein products.
	D.J.L. Luck and E. Reich	Isolate mitochondrial DNA from *Neurospora.*
1965	R.W. Holley and coworkers	Complete base sequence of alanine transfer RNA of yeast.
	S. Brenner, A.O.W. Stretton and S. Kaplan	Deduce that UAG and UAA signal polypeptide chain termination.
	H. Harris and J.F. Watkins	Fusion of man and mouse cells using sendai virus.

	DuPraw	'Folded fiber model' of eukaryotic chromosome organisation.
1966	B. Weiss and C.C. Richardson	Isolate DNA ligase.
	F.H.C. Crick	Wobble hypothesis of base-pairing between *m*RNA codon and tRNA anticodon.
	J. Adams and M. Cappecchi	N-formylmethionyl-tRNA functions as the initiator of polypeptide chain formation.
	W. Gilbert and B. Müller - Hill	*lac* operon repressor is a protein.
	M. Ptashne	Phage λ-repressor is a protein, and binds directly to λ DNA.
1966	M. Waring and R.J. Britten	Repetitive sequences in vertebrate DNA.
	V.A. Mckusick	Publishes '*Mendelian Inheritance in Man*'; records Ca. 1,500 genetic disorders.
1967	R. Sager and Z. Ramanis	Biparental inheritance of plasmagenes in *Chlamydomonas;* caused by irradiation of cells with UV before mating.
	H.G. Khorana and coworkers	Use synthetic polynucleotides of known repeating di-or trinucleotide sequences to resolve the genetic code.
1968	R.T. Okazaki and coworkers	Newly replicated DNA contains many fragments, now called 'Okazaki fragments'.
	W. Gilbert and D. Dressler	'Rolling circle model' of DNA replication.
	M. Kimura	'The neutral gene theory' of molecular evolution.
	H.O. Smith, K.W. Wilcox and T.J. Kelley	Isolate and characterise the first restriction endonuclease, *Hin*dIII.
	J.A. Huberman and A.D. Riggs	Mammalian chromosome contains serially arranged replicons, each being ea. 30 μm long.
	S. Wright	Publishes '*Evolution and the Genetics of Populations*' Vol. I; four volume series; Vol-4 published in 1978.
1969	R. Burgess and coworkers	Isolate and identify the σ factor from RNA polymerase.
1970	H.G. Khorana and 12 coworkers	Total synthesis of gene for an alanine tRNA from yeast.
	D. Baltimore and H.M. Temin	Reverse transcriptase in two oncogenic RNA viruses.

	R. Sager and Z. Ramains	First map of plasmagenes; 8 genes mapped in chloroplast DNA of *Chlamydomonas.*
	Ohno	Proposed that new genes originate due to duplication followed by continued mutation in one of the two copies of the duplicated gene.
1971	K. Dana and D. Nathans	First restriction map (of SV40 genome)
1972	S.J. Singer and G.L. Nicholson	Fluid-mosaic model of cell membranes.
	R. Silber, V.G. Malathi and J. Hurwitz	Discover RNA ligase.
	D.A. Jackson, R.H. Symons and P. Berg	First recombinant DNA molecule; joined DNA of SV40 and phage λ.
	P.S. Carlson, H.H. Smith and R.D. Dearing	The first somatic hybrid plant in *Nicotiana;* produced by protoplast fusion.
1973	R. Kavenoff and B.H. Zimm	A chromosome contains a single long molecule of DNA; used viscoelastic method for measuring weights of DNA molecules isolated from different *Drosophila* species.
	W. Fiers, W.M. Jou, G. Haegerman and M.Ysebaert	First to sequence a gene encoding a protein; the coat-protein gene of the male-specific phage MS2.
1974	J. Shine and L. Dalgarno	Discover 'Shine-Dalgarno' sequence at the 3'-end of *E. coli* 16*S* rRNA, which is complementary to the ribosome binding sites of mRNAs.
	I. Zaenen and coworkers	Ti plasmid of *Agrobacterium tumefaciens.*
	R.D. Kornberg	Proposes the 'nucleosome-solenoid model of chromatin'; universally accepted.
	G.L. Stebbins	Publishes *'Flowering Plants, Evolution Above the Species Level'*.
1975	G. Köhler and C. Milstein	Develop the hybridoma technology for monoclonal antibody production.
	M. Grunstein and D.S. Hogness	Colony hybridization technique.
	D. Pribnow	A model for prokaryotic promoter structure.
	E.M. Southern	Southern blotting and hybridization technique.

	F. Sanger and A.R. Coulson	Enzymatic method of DNA sequencing.
1976	M.F. Gellert and coworkers	Discover DNA gyrase and its functions.
	B.G. Burrel and coworkers	Overlapping genes in phage ϕ *X* 174.
	—	N IH, U.S.A. issues guidelines regulating recombinant DNA research.
1977	C. Jacq, J.R. Miller and G.G. Brownlee	Pseudogenes within the 5*S* DNA cluster of *Xenopus laevis.*
	J.C. Alwine, D.J. Kemp and G.R. Stark	Develop the 'northern blotting' technique.
	F. Sanger and cowokers	Base sequence of the complete genome of ϕ *X* 174.
	M. Leffak, R. Grainger and H. Weintraub	During DNA replication, 'old' histone occtamers of nucleosomes remain intact, and 'new' octamers consist of proteins synthesized during DNA replication.
	A.M. Maxam and W. Gilbert	'Chemical method' of DNA sequencing.
	D.S. Hogness and coworkers	Report introns in some 28*S* rRNA genes of *D. melanogaster.*
	A. Jeffres and R.A. Flawell	Introns in rabbit beta - globin gene.
	P. Chambdon and coworkers	Introns in chicken ovalbumin gene.
	J. Collins and B. Holm	Develop cosmids.
1978	R.M. Schwartz and M.O. Dayhoff	Show from extensive protein and nucleic acid sequence data that protoeukaryotic organisms entered into symbiosis with mitochondria and chloroplasts about 2 and 1 billion years, respectively, ago.
	W. Gilbert	Coins the terms '*introns*' and '*exons*'.
	T. Maniatis and Others	First 'genomic library' for gene isolation.
	Y.W. Kan and A.M. Dozy	Use RFLP for prenatal diagnosis of sickle cell anemia.
	C.A. Hutchison and coworkers	Site- directed mutagenesis in isolated DNA.
	E.H. Blackburn and J.G. Gall	*Tetrahymena pyriformis* telomere consists of 30 to 70 repeats of the sequence AACCCC (complementary sequence, TTGGGG).

1979	D.V. Goeddel and coworkers	Using recombinant DNA technology, prepare a gene construct encoding human growth hormone (hGH) and express it in *E. coli*; a hGH like polypeptide produced.
	M.R. Lerner and J.A. Steitz	Discover small nuclear RNAs (snurps).
1980	J.W. Gordon and coworkers	First transgenic mice.
	A.R. Templeton	'Founder principle' for speciation.
1981	L. Margulis	Publishes '*Symbiosis in Cell Evolution*'.
	J.D. Kemp and T.H. Hall	Transfer phaseolin gene from beans to sunflower using Ti plasmid vector.
	T.R. Cech, A. J. Zaug and P.J. Grabowski	Self-splicing rRNA in *Tetrahymena thermophila.*
	S. Anderson and 13 others	Complete base sequence and organization of human mitochondrial genome.
1982	Eli Lilly International corporation	First to market a drug produced by recombinant DNA technology; insulin marketed as 'Humulin'.
1983	E.A. Miele, D.R. Mills and F.R. Kramer	First recombinant RNA molecule.
1985	C. W. Greider and E.H. Blackburn	Isolated telomerase from *Tetrahymena pyriformis.*
	R.K. Saiki, K.B. Mullis and five others	Polymerase chain reaction to amplify β-globin gene fragments *in virto.*
1986	M.C. Shih, G. Lazar and H.M. Goodman	In higher plants, nuclear genes encoding chloroplast glyceraldehyde - 3 - phosphate dehydrogenase are derived from the genomes of symboints, which gave rise to chloroplasts.
1987	(P. Ledev and T. Stewart)*	Harvard University receives the patent for oncomouse, first U.S. patent for a genetically altered animal.
1990	Sinclair	*SRY* in human *Y* chromosome; it encodes a 'testes determining factor'; presumably a transcription factor.

*The scientists who developed the oncomouse.

Appendix-II

A CHRONOLOGICAL LIST OF THE SCIENTISTS AWARDED NOBEL PRIZE FOR DISCOVERIES IN GENETICS AND RELATED AREAS

Year of award	*Name of scientist(s)*	*Work for which awarded*
1901	E. von Behring	Studies on antiserum therapy
1930	K. Landsteiner	Studies in immunology
1933	T.H. Morgan	*Drosophila* genetics, including sex-linkage and linkage, *i.e.*, development of theory of gene.
1935	H. Spemann	Studies on embryonic induction.
1945	A. Fleming, E.B, Chain, and H.W. Florey	Discovery, purification and chemical characterization of penicillin.
1946	H.J. Muller	Mutagenic action of X-rays in *D. melanogaster.*
	J.B. Sumner	Crystallization of enzymes.
	W.M. stanley	Purification and chemical characterization of viruses.
1952	A.J.P. Martin and R.L.M. Synge	In chemistry; for chromatography technique.
1957	A. Todd	Studies on the structure of nucleosides and nucleotides.
1958	G.W. Beadle, E.L. Tatuam and	*Neurospora* biochemical genetics (Beadle, Tatum) and bacterial conjugation (Tatum, Lederberg).
	J. Lederberg	
	F. Sanger	Studies on protein chemistry.
1959	S. Ochoa and A. Kornberg	*In vitro* synthesis of DNA (Kornberg) and RNA (Ochoa).
1960	P.B. Medawar and F.M. Burnet	Immunological tolerance.
1962	J.D. Watson, F.H.C. Crick, and M.H.F. Wilkins	The double helix model of DNA (studies on DNA structure).
1962	M.F. Perutz and J.C. Kendrew	Studies on the structures of hemoglobin and myoglobin.
1965	F. Jacob, J. Monod, and A. Lwoff	Operon concept of gene regulation; microbial genetics (A. Lwoff).
1966	P. Rous	Studies on oncogenic viruses.

1968	M.W. Nirenberg, Hargovind Khorana and R.W. Holley	Deciphering of genetic code (Nirenberg, Khorana, Holly) and *in vitro* synthesis of DNA (Khorana).
1969	M. Delbrück, S.E. Luria and A.D. Hershey	Contributions to viral genetics.
1972	G.M. Edelman and R.R. Porter	Studies on chemical structure of antibodies.
1974	A. Claude, C. de Duve and G. Palade	Contributions to cell biology.
1975	R. Dulbecco, H. Temin and D. Battimore	Studies on oncogenic viruses.
1977	R.S. Yalow	Development of radioimmunoassay procedure.
1978	H.O. Smith, D. Nathans and W. Arber	Development of techniques utilizing restriction endonucleases for the study of genetic organization.
1980	P. Berg, W. Gilbert and F. Sanger	In chemistry: for experimental manipulation of DNA, including production of recombinant DNA.
	G.D. Snell, J. Dausset and B. Benacerraf	In physiology and Medicine; for contributions to immunogenetics.
1982	A. Klug	Analysis of crystalline structures of biological importance, especially virus particles, tRNA and nucleosomes.
1983	B. McClintock	Discovery of transposable elements (*Ac* and *Ds*) in maize.
1984	N.K. Jerne, G. Köhler and C. Milstein	Contributions to immunology, including hybridoma technology.
	R.B. Merrifield	In chemistry; automated peptide synthesis.
1985	M.S. Brown and J.L. Goldstein	Identification of low density lipoprotein receptor pathway; familial hypercholesterolemia is a genetic defect in this pathway.
1986	E. Ruska	Designing of first electron microscope.
1987	S. Tonegowa	Genetic mechanism that generates antibody diversity.
1989	J.M. Bishop and H.E. Varmus	In Medicine; studies on oncogenes of retroviruses.
	T.R. Cech and S. Altman	In chemistry; enzymatic RNAs (= ribozymes).

1992	E.H. Fischer and E.G. Krebs	Discoveries concerning reversible protein phosphorylation as a biological regulatory mechanism.
	K.B. Mullis	In Chemistry; invention of the PCR method.
	M. Smith	In Chemistry; fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies.
1993	R.J. Roberts P.A. Sharp	For the discovery of split genes.
1994	A.G. Gilman and Martin Rodbell	Discovery of G-proteins and the role of these proteins in signal transduction in cells.
1995	E.B. Lewis, C. Nüsslein—Volhard and E.F. Wieschaus	The discoveries concerning the genetic control of early embryonic development.
1997	S.B. Prusiner	The discovery of prions—a new biological principle of infection.
1999	G. Blobel	The discovery that proteins have intrinsic signals that govern their transport and localization in the cell.
2001	L.H. Hartwell, R.T. Hunt and Sir P.M. Nurse	Discoveries of the key regulators of the cell cycle.
2002	Sidney Brenner, H.R. Horvitz and J.E. Sulston	Discoveries concerning genetic regulation of organ development and programmed cell death.
	J.B. Fenn	In Chemistry; development of methods for identification and structure analyses of biological macromolecules.

Appendix-III

TERMS COINED BY

Term	*Coined by*
A chromosome	L.F. Randolph (1928)
Allelomorph	W. Bateson (ca. 1902—1909)
Bacteriophage	F. d' Herelle (1917)
B chromosome	L.F. Randolph (1928)
Chromatin	W. Flemming (1879)
Chromosome	W. Waldeyer (1888)
Cistron	S. Benzer (1955)
Coupling phase	W. Bateson and R.C. Punnet (1906)
Dosage compensation	H.J. Muller (1948)
Endoplasmic reticulum	K.R. Porter (1953)
Enzyme	W. Kuhne (1878)
Epistatic gene	W. Bateson (Ca. 1902—1909)
Eukaryote	E. Chatton (1937)
Exon	W. Gilbert (1978)
F_1 and F_2	W. Bateson (Ca. 1902—1909)
Gene	W. Johannsen (1909)
Genetics	W. Bateson (1905)
Genotype	W. Johannsen (1909)
Heterochromatin	E. Heitz (1928)
Heterozygote	W. Bateson (Ca. 1902—1909)
Histone	A. Kossel (Ca. 1890)
Homozygote	W. Bateson (Ca. 1902—1909)
Intersex	R.B. Goldsmidt (1915)
Intron	W. Gilbert (1978)
Lysosome	C. de Duve *et al.* (1955)
Meiosis	Farmer (1905)
Mitochondria	C. Benda (1890)
Mitosis	W. Flemming (1882)
Mutation	H. de Vries (1901)
Muton	S. Benzer (1955)

Nucleic acid	Altman (1889)
Phenotype	W. Johannsen (1909)
Polygenes	K. Mather (1941)
Prokaryote	E. Chatton (1937)
Protein	G. Mulder (1838)
Recon	S. Benzer (1955)
Repulsion phase	W. Bateson and R.C. Punnet (1906)
Terminalization (of chiasma)	C.D. Darlington (1931)
Transition	E. Freese (1959)
Transversion	E. Freese (1959)
Virus	M.W. Beijerinck (1899)
X Chromosome	Wilson and Stevens (1905)
Y Chromosome	Wilson and Stevens (1908)

Appendix-IV

MODELS, THEORIES AND HYPOTHESES

	Theory/ hypothesis	*Proposed by*
1.	Cell lineage theory	R. Virchow (1855)
2.	Cell theory	M.J. Schleiden and T. Schwann (1838-39)
3.	Chiasmatype or one-plane theory	F.A. Janssens (1909, 1924)
4.	Classical or two-plane theory	Sharp (1934)
5.	Copy-choice theory	J. Lederberg (1955)
6.	Double helix model of DNA	J.D. Watson and F.H.C. Crick (1953)
7.	Evolution through natural selection	C. Darwin and A.R. Wallace
8.	First genetic map (in *Drosophila*)	A.H. Sturtevant (1913)
9.	Fluid mosaic model of cell membranes	S.J. Singer and G.L. Nicholson (1972)
10.	Folded-fiber model of eukaryotic chromosome	Dupraw (1965)
11.	Founder principle of speciation	A.R. Templeton (1980)
12.	Genic balance theory of sex determination	C.B. Bridges (1925)
13.	Hardy-Weinberg law	G.H. Hardy and W. Weinberg (1908)
14.	Hybrid DNA models of recombination	Whitehouse (1963), Holiday (1964)
15.	Inheritance of acquired characters	J.B. de Monet Lamarck (1809)
16.	Law of independent assortment	G.J. Mendel (1865)
17.	Law of segregation	G.J. Mendel (1865)
18.	Model for prokaryotic promoter	D. Pribnow (1975)
19.	Multiple factor hypothesis	H. Nilsson-Ehle (1909)
20.	Nucleosome-solenoid model of chromatin fibers	R.D. Kornberg and Thomas (1974)
21.	One-gene-one enzyme hypothesis	G.W. Beadle and E.L. Tatum (1941)
22.	One-gene-one-polypeptide hypothesis	V.M. Ingram (1957)
23.	Operon concept of gene regulation	F. Jacob and J. Monod (1961)
24.	Pureline theory	W. Johannsen (1909)
25.	Replicon model	F. Jacob and S. Brenner (1963)
26.	Rolling circle model of DNA replication	W. Gilbert and D. Dressler (1968)
27.	The neutral gene theory of molecular evolution	M. Kimura (1968)
28.	Wobble hypothesis	F.H.C. Crick (1966)

The book Genetics has been designed to accommodate the recent developments in this rapidly growing subject. This book has a total of 62 chapters covering historical developments, Laws of Mendel and other aspects of transmission genetics; cell division and cell cycle regulation; recombination (including molecular basis); genome mapping; genetic material, its organization, replication and repair; gene action (transcription, translation, RNA processing) and its regulation; developmental genetics; recombinant DNA technology, human genetics and cancer, and biosafety. There is a marked emphasis on molecular aspects of genetics and on recombinant DNA technology; this reflects the increasing relevance of these aspects in modern biology.

The science of genetics is developing at a very rapid pace, and it is very difficult for a text book to keep pace with these developments. In addition, the considerations of space force many deletions and condensations that may be to the detriment of a fluent treatment. The author has attempted a balanced coverage keeping in view the needs of modern revised syllabi of genetics, and the trends of recent developments in the subject.

It is hoped that ***'Genetics'*** will provide the post-graduate students a basic study material on the various aspects of genetics. The treatment of subject retains the basic approach of the earlier book and attempts to explain the various concepts and phenomena in a simple, clear and easy-to-follow discussion.

The new edition (3rd edition) of ***'Fundamentals of Genetics'*** has been toned down and condensed with a view to adapt it to the requirements of undergraduate students of Indian Universities. The contents, however, have been revised and expanded to include recent developments in the subject and the changing needs of the students.

ISBN 81-272-1331-4

Rs. 240.00

KALYANI PUBLISHERS